Relativistic Field Theory for Microwave Engineers

MATTHEW A. MORGAN

ARTECH BOOKS

Relativistic Field Theory for Microwave Engineers

For a complete listing of titles in the
Artech House Microwave Series,
turn to the back of this book.

Relativistic Field Theory for Microwave Engineers

Matthew A. Morgan

ARTECH
HOUSE

BOSTON | LONDON
artechhouse.com

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the U.S. Library of Congress.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library.

Cover design by Joi Garron

ISBN 13: 978-1-68569-067-0

685 Canton Street
Norwood, MA 02062

10 9 8 7 6 5 4 3 2 1

To Zhenya
Ты мое сокровище

Contents

Preface

"The magnetic field is really just a relativistic correction for the electric field."

Thus remarked my undergraduate advisor to one of his graduate research assistants in the hallway, some lifetimes ago for me. The comment was not intended for my ears, and as an undergraduate at the time, I knew that I was not prepared to understand it, but still, the thought fascinated me. The needs of my career demanded that I spent the next two decades practicing electromagnetics in the more normative fashion for my profession, wherein the electric and magnetic fields are treated as dualistic, inexorably linked, nevertheless distinct mathematical quantities, but this small kernel of truth has always stuck fast in a private corner of my mind, wanting exploration. In this book, that seed planted by my mentor those many years ago has finally taken root.

It is a great pity, I think, that so few students and professionals trained in the laws of applied electromagnetics understand the intimate linkage between those laws to which they have dedicated their lives and the principles of special relativity. Most of us are taught — if we learn about special relativity at all — that Einstein conceived of special relativity simply by imagining what it would be like to travel alongside a beam of light, concluding that nothing could move faster and that therefore the definitions of space and time needed revision. Even as an undergraduate, I remember wondering what makes light so special? Why not the speed of sound, or the speed of particle radiation? It surely isn't that light is fast. Many things are fast, from our human perspective, which in itself is pretty meaningless on a cosmic scale. Why not hold up something incredibly slow and then say that nothing can move slower? No, there had to be something about light, something unique to warrant granting it that special place as the universal governor of relative speed. In fact, Einstein had great reasons for holding light — electromagnetic waves — in such high regard. Those reasons are woven, directly and inextricably, into Maxwell's equations. This book will explain why.

I also find it curious that relativity was developed prior to quantum mechanics[1] (two modern theories that we still struggle to reconcile), but the latter has been applied to a vast array of practical technologies, while the former has languished substantially unused, save for the one common example of the global positioning systems, which account for the slight relativistic time dilation observed in the atomic clocks on board satellites in orbit (and even in that case, relativity is not so much exploited as it is corrected for, as the annoying friction of a wheel bearing is suppressed with grease). The science of relativity remains a virtual curiosity to most, practiced only theoretically by astrophysicists and discussed in fanciful thought experiments by the academic elite.

Theoreticians might argue that relativity only applies at such large energies, with such rapidly moving bodies, or over such great distances that it can have no practical bearing on the daily human experience and that controlling the conditions needed to exploit it are beyond our capability. To me, such a statement is nothing more than an excuse for our lack of foresight. One might just as well have argued that the microscopic scales and low energies at which the interactions of single particles elucidate quantum mechanical effects are equally beyond the scope of normal human experience, and that controlling the conditions on the subatomic scale necessary to leverage quantum mechanics should be equally difficult for us — yet we have accomplished exactly that. The tablet computer you are probably now holding could well be characterized as a quantum-mechanical machine.

It is my belief that our failure to apply relativity lies not in its practical inaccessibility, but in the exclusivity with which it is taught. Theoretical physicists learn deeply about relativity and then apply it to observations of cosmological events, but engineers are rarely exposed to it beyond a brief mention in undergraduate physics and almost never in the context of their own field of expertise. With all due respect to theoretical physicists, many of whom I work with on a daily basis, it is plainly apparent to me that scientists and engineers think differently. We need to put relativity in the hands of creative engineers if we ever wish to harness its power.

As we will see in this book, the reality is that we are already exploiting relativistic effects, though precious few of us ever realize it. As my advisor said, magnetic fields are themselves a relativistic effect, generated by moving electric charges and influencing others that also move. The electrons in our circuits do not generally move even close to the speed of light, but the magnetic effect is strong, because the fundamental electrodynamic force is strong. Relativity is like the weak modulation applied to a strong carrier. This will be the key to our unlocking relativity at low energies.

---

1 In fact, today we consider special relativity as a branch of classical physics.

To be clear, I do not claim that any new physics will be exposed here. All of the results that will be derived could also be explained entirely within the context of classical electromagnetics, with which I assume the reader is already familiar. I have personally seen and experienced, however, the value that a simple change in perspective can have in allowing one to come up with new approaches to problems that would otherwise defy our intuition.

For reasons that I hope I have explained, this book is targeted for engineering students and practicing engineering professionals, not theoretical physicists (who already have plenty of books on the subject available to them). As such, I wish to justify what will likely be a fairly controversial choice I have made: to present this work in the International System of Units (SI), as opposed to the Gaussian-Centimeter-Gram-Second (CGS) unit system, which is touted as having advantages for this particular subject matter.

The Gaussian system recasts the physical laws of electromagnetics in such a way that the electric and magnetic fields each have the same units, a marked convenience when one is attempting to describe them as merely two coordinate axes of the same spacetime plane. (One reference I have relied on heavily likens the SI units to measuring north-south distances in meters and east-west distances in feet.) In my view, however, the relative obscurity of the Gaussian system makes it a serious roadblock to those trained in SI from understanding, or even accepting, the results derived from them. Despite the insistence of Gaussian proponents that such things are natural, I cannot be comfortable measuring capacitance in centimeters, or resistivity in seconds. Something deep in my gut is telling me that what you are calling a resistance is actually something else, perhaps related, but certainly not resistance as I understand it. By adopting Gaussian units, those of us trained in SI are required to suspend our disbelief at the outset, when the supposedly fundamental laws are rewritten in these new terms. How then, can we be expected to trust the final result, which so blatantly defies our experience?

Better to start on firm and familiar ground, and see where it takes us, however inconvenient the notation may be initially. Only then, once we are satisfied as to the correctness of the result, can we adapt our thinking to a new system of measurement. If indeed we find that the electric and magnetic fields are two dimensions of the same physical quantity, one measured in meters and the other in feet (proverbially speaking), then so much more will be our appreciation of that fact, given the unlikelihood of its foundations. Notational convenience is extremely important — in consideration of which I have included Appendix A, which reproduces the key results of classical electromagnetics in Gaussian-CGS form — but first we must understand why.

Acknowledgments

I am grateful to my once undergraduate advisor and now friend and colleague, Dr. Bobby Weikle, for lighting the spark (without realizing it) that drove me to learn more about this subject and ultimately to write this book. Bobby is well aware of my previous literary efforts, but I have kept this project somewhat closer to the vest, so perhaps I will surprise him this time.

I would also like to thank Alex Arsenovic of 810 Labs (http://810lab.com) for opening my eyes to the elegance of spacetime algebra, the perfect mathematical language with which to illuminate the concepts contained in these pages. Were it not for his consultations, imparted over barbecue and burritos, my readers would be forced to learn about this difficult subject through the comparatively hazy lens of tensor analysis. I have decided to allow one chapter on the subject of tensors (Chapter 4) to remain in the finished product, in part due to its historical significance and prevalence in the literature, but also so that you may understand what a genuine tragedy it would have been had Alex not shown me a better way.

Chapter 1

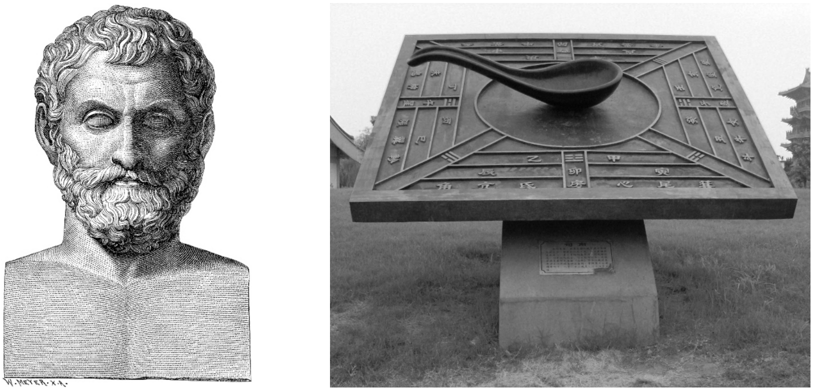
Classical Electromagnetics

The principles of special relativity, though widely accepted as fact in the scientific community, are in such stark contradiction with ordinary human experience that they are notoriously difficult to accept at first exposure. Since these principles, as I hope to show in this book, follow directly from classical electromagnetic theory, it is deemed a worthwhile exercise to first review the experimental origins of that theory — in essence, so that we may understand how it is we know what we know. In this way, the intuitive leap that will later be required of the reader as we pass from the familiar to the fantastical need not be based on blind faith alone, but instead on the confidence of a firm foundation laid down by sound, scientific methodology.

1.1 EARLY CONCEPTS IN ELECTRICITY AND MAGNETISM

Although early scholars lacked the tools, both physical and mathematical, to properly explore and understand electromagnetic forces, clues of their existence were nevertheless in evidence throughout the natural world, for those empowered with the intellectual curiosity to notice them.

It has been known since antiquity, for example, that certain materials acquire an electrostatic charge when rubbed against other materials and then separated, subsequently exhibiting an attractive force toward one another and other light objects (e.g., feathers) — a form of contact electrification known as the *triboelectric effect* [1]. Amber, for example, becomes statically charged when rubbed against wool, as does glass when rubbed against silk, and hard rubber against fur. The word triboelectric, in fact, derives from the Greek tribo-, for rub, and ēlektron, for amber. This property of materials was first recorded circa 500 BC by Thales of Miletus,



(a) (b)

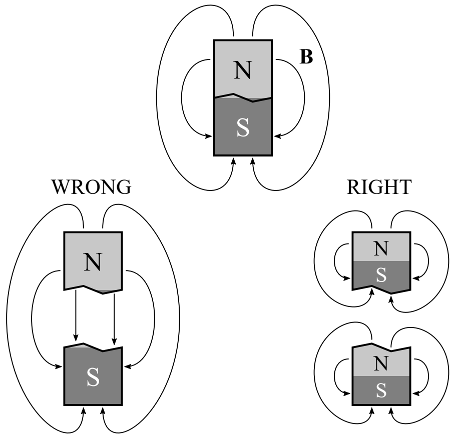
Figure 1.1 (a) Thales of Miletus, who wrote some of the earliest known texts describing the triboelectric effect and the magnetic properties of lodestone. (b) An ancient Chinese compass known as a *sinan* (also *Zhinan*), or south-pointing lodestone spoon/ladle.

shown in Figure 1.1(a), a Greek philosopher and mathematician who is credited for breaking from the traditions of his contemporaries by attempting to explain the natural world in terms of testable hypotheses instead of mythology, in anticipation of the modern scientific method [2].

Thales also provided some of the earliest known references to the magnetic properties of *lodestone* — a magnetized form of the mineral magnetite, Fe_3O_4. It has been suggested that the word *magnet* derives from Magnesia, a Greek city (settled by the Magnetes) in what is now present-day Turkey, where lodestones were found [3]. Lodestone was also referenced by the Chinese in the fourth-century BC, who shaped it into a spoon, which, balanced on a smooth base plate, would pivot to the south as a type of compass, as seen in Figure 1.1(b). There is further some evidence to suggest that lodestone was known to the Olmecs, a Mesoamerican civilization, some thousand years prior.

Thales incorrectly asserted that amber became magnetized as a consequence of friction, unlike lodestone which needed no friction. Although his conclusions were incorrect, he nevertheless anticipated the linkage between electric and magnetic phenomenon, which would not be proven decisively for over 2,000 years.

(a)



(b)

Figure 1.2 (a) William Gilbert, author of *De Magnete*, 1600, who also coined the term electricity. (b) Illustration that permanent magnets, having north and south poles, if split in two, do not divide into two magnetic monopoles.

1.2 ADVANCEMENT THROUGH EXPERIMENTATION

From these humble beginnings, our understanding of electromagnetic forces increased over the next 2,000 years through a process of meticulous observation and experimentation to become one of the most accurate and successful disciplines in modern science and engineering. As is always the case, pure scientific reasoning and technological innovation went hand in hand, as greater understanding of fundamental physical laws alternatively required and then led to the creation of ever more precise equipment with which to make careful measurements [4, 5]. A detailed account of every step along the way is beyond our scope. However, highlights of the most critical experiments and breakthroughs that underpin electromagnetic field theory will be summarized in the following sections.

1.2.1 Absence of Magnetic Monopoles

In 1600, an English physicist named William Gilbert, seen in Figure 1.2(a), published a landmark scientific work called *De Magnete, Magneticisque Corporibus, et de Magno Magnete Tellure* (On the Magnet and Magnetic Bodies, and on the

Great Magnet the Earth) [6]. Gilbert recognized the importance of the fact that a magnet, having a north pole and a south pole, when cut in two yields not two isolated magnetic monopoles, but rather two new magnets each having a north and south pole of their own, as seen in Figure 1.2(b).

It would be over two centuries before this basic principle would be described in terms of the lines of force known as magnetic flux density, $\mathbf{B}$, and another half a century after that before the vector-calculus notation existed to write it down in the form that we know today,

$$\boldsymbol{\nabla} \cdot \mathbf{B} = 0 \tag{1.1}$$

where the *del operator*, $\boldsymbol{\nabla}$, in Cartesian coordinates is given by [7]

$$\boldsymbol{\nabla} = \mathbf{x}\frac{\partial}{\partial x} + \mathbf{y}\frac{\partial}{\partial y} + \mathbf{z}\frac{\partial}{\partial z} \tag{1.2}$$

and thus (1.1) becomes

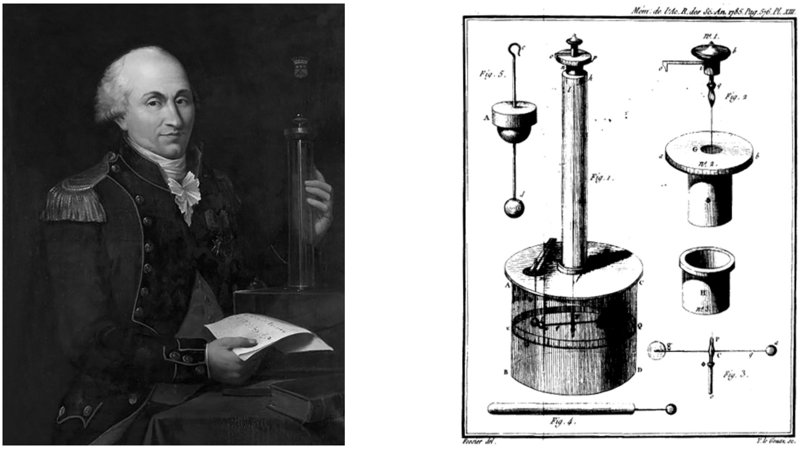
$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0 \tag{1.3}$$

This is, of course, the first of four fundamental laws of classical electromagnetics known as *Maxwell's equations*.[1] It states that the magnetic flux density has no divergence; in plainer terms, it does not spread apart. Its field lines therefore do not emanate from anywhere (the location of which would be called magnetic monopoles); rather, they form closed loops.

1.2.2 Electrostatic Force and Gauss's Law

By the mid-1700s, various researchers around the world were hypothesizing and conducting experiments to confirm that the force of attraction or repulsion between static charges varied as the square of the distance between them, a mathematical relationship the force of gravity was by then known to exhibit [8]. Daniel Bernoulli of Switzerland and Alessandro Volta of Italy both measured the forces exerted by the charged plates of a capacitor upon one another. John Robison, a Scottish physicist,

1 Like most researchers, I refer to the set of four multivariable differential equations known as *Ampère's law*, *Faraday's law*, and *Gauss's laws* for electricity and magnetism as Maxwell's equations. In fact, Maxwell used a set of 20 scalar equations that encompassed these as well as numerous other relationships between the auxiliary fields, potentials, electrical resistance, and charge continuity. Not all were mutually independent (nor did he claim that they were), but they were convenient for use in his time. Much of the material in this book is dedicated to condensing Maxwell's original equations, including the critical four that bear his name today, into even more efficient and elegant forms.



(a) (b)

Figure 1.3 (a) Charles-Augustin de Coulomb, who derived the law of electrostatic force using (b) a torsion balance.

concluded in 1769 from his measurements on spheres carrying charges of the same polarity, that they repelled each other with a force inversely proportional to the distance raised to the 2.06 power.[2]

Final credit for the inverse-square law, however, is generally given to Charles-Augustin de Coulomb, seen in Figure 1.3(a), who in 1785 definitively published the law that now bears his name,

$$F_e = \frac{q_1 q_2}{4\pi\varepsilon_0 r^2} \tag{1.4}$$

where, in present-day notation, q_1 and q_2 are the two charges, ε_0 is the vacuum permittivity ($\approx 8.8541878176 \times 10^{12}$ F/m), and r is the distance between them. The electric force, F_e, is repulsive as written when both charges have the same sign, attractive when the signs differ, and is directed along the line between them.

Coulomb performed his measurements using a *torsion balance* (Figure 1.3(b)), a device for measuring very weak forces that he himself devised in 1777. It comprised an insulating bar suspended by silk thread as the torsion spring. At one end of the bar was a conducting sphere to which a charge is applied. Another

2 One should not be too swift to disparage the obvious error in the second decimal place of Robison's result. Empirical data must always precede theory, and all measurements, no matter how careful, are subject to reasonable margins of error. Robison should be commended for letting his data speak for itself, rather than substituting what he surely must have felt was the preferred answer, 2, in its place.

sphere having the same charge was brought near, the repulsive force causing the silk thread to twist. By measuring the angle of deflection, and knowing how much force was required to twist the thread by the same angle, Coulomb was able to carefully measure the electrostatic force under varying conditions and arrive at the law.

In the modern concept, the electric field vector, $\mathbf{E}$, is the directed line of force imparted by a known, existing charge upon a hypothetical test charge, q, located anywhere in space, or

$$\mathbf{F}_e = q\mathbf{E} \tag{1.5}$$

Given the additional fact (born out by Coulomb's and others' measurements) that the electric fields from multiple charges obey superposition, it is a fairly simple matter to convert Coulomb's law into Gauss's law for the electric field generated by an arbitrary distribution of charge density, ρ. First, let us write the divergence as an integral of the vector field normal to a surface enclosing an infinitesimally small volume of space,[3]

$$\boldsymbol{\nabla} \cdot \mathbf{E} = \lim_{\Delta V \to 0} \frac{1}{\Delta V} \oiint \mathbf{E} \cdot d\mathbf{S} \tag{1.6a}$$

$$= \lim_{\Delta V \to 0} \frac{1}{\Delta V} \int_0^{2\pi} \int_0^{\pi} \frac{q}{4\pi\varepsilon_0 r^2} r^2 \sin\theta d\theta d\phi = \lim_{\Delta V \to 0} \frac{q}{\varepsilon_0 \Delta V} = \frac{\rho}{\varepsilon_0} \tag{1.6b}$$

$$\therefore \boldsymbol{\nabla} \cdot \mathbf{D} = \rho \tag{1.6c}$$

where $\mathbf{D} = \varepsilon_0 \mathbf{E}$ is called the *electric flux density* or *displacement*.

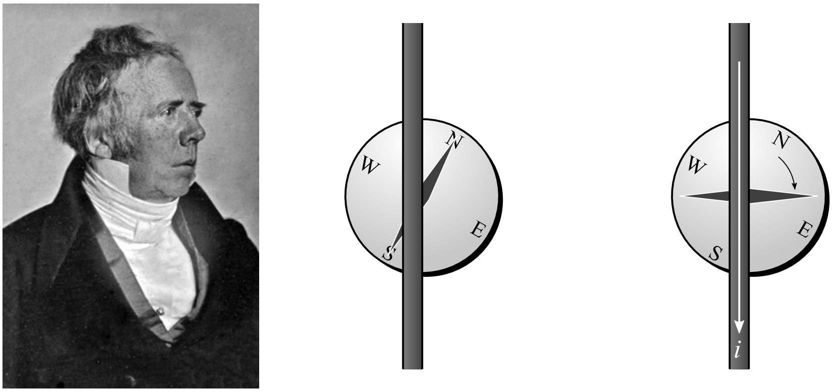
1.2.3 Lorentz Force and Ampère's Law

Coulomb experimented with magnets also, measuring the forces of attraction and repulsion between the north and south poles, but the linkage between electricity and magnetism was still not known in his time. Like most of his contemporaries, he believed that these forces were the independent consequences of two distinct fluids.

Evidence to the contrary would be discovered quite by accident[4] in 1820 by the Danish physicist, Hans Christian Ørsted (commonly written "Oersted" in

3 It will generally be assumed that the reader is familiar with vector differential forms, integral theorems, and related concepts. A brief summary of the key mathematical elements is given in Appendix B.

4 The accidental nature of the discovery is somewhat disputed [9]. While it is evident that Ørsted (and others) were by that time actively seeking a linkage between electricity and magnetism, Ørsted was convinced by reasons of symmetry that the magnetic field should be longitudinal, or parallel to the wire, and designed his experiments accordingly. He noticed the transverse magnetic effect by chance while concluding another failed attempt to demonstrate it longitudinally.



(a) (b) (c)

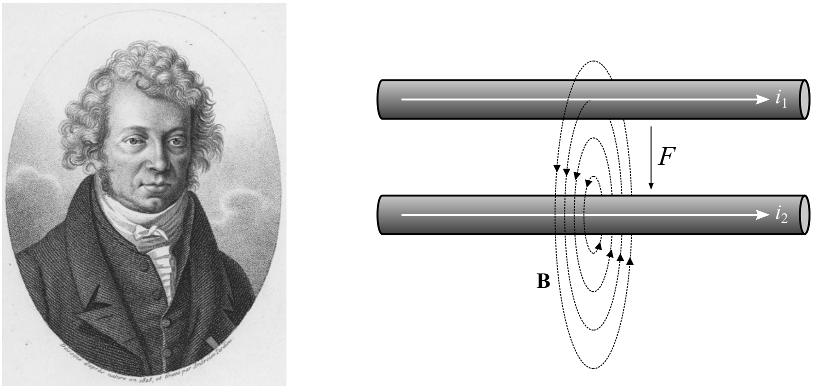
Figure 1.4 (a) Hans Christian Ørsted, Danish physicist and discoverer of the magnetic field created by an electrical current. (b) Compass and wire with the current off. (c) Compass and wire with the current on.

English), seen in Figure 1.4(a). While giving a lecture and demonstration, Ørsted noticed a compass needle deflecting from magnetic north whenever he switched on the current in a nearby wire [10], illustrated in Figure 1.4(b) and (c). Following up on these observations, Ørsted conducted a number of experiments that ultimately led him to publish the result that a current flowing through a wire generates a circulating magnetic flux density around it.

Later that same year, a French physicist named André-Marie Ampère, shown in Figure 1.5(a), learned of Ørsted's findings when his friend, François Arago, presented them to the French Academy of Sciences. Ampère conducted experiments of his own to better quantify and apply a mathematical structure to Ørsted's discovery [11]. He eventually arrived at what would become known as *Ampère's force law*, describing the force of attraction between two parallel, current-carrying wires,

$$F_m = \frac{\mu_0 i_1 i_2}{2\pi r} L \tag{1.7}$$

where i_1 and i_2 are the signed currents in the two wires, μ_0 is the vacuum permeability ($4\pi \times 10^{-7}$ H/m), r is the distance between the two wires, and L is their length. The force is attractive when both currents are oriented in the same direction and repulsive when they flow in opposite directions.



(a) (b)

Figure 1.5 (a) André-Marie Ampère, French physicist who described the force between parallel, current-carrying wires. (b) Illustration of the magnetic field surrounding one wire and Ampère's force law drawing the other wire to it. Similarly, the second wire would generate its own magnetic field drawing the first wire to it (not shown).

Recognizing today that currents in wires are nothing more than electric charges in motion, we may rewrite this force for a single unit of that charge, q, in vector form

$$\mathbf{F}_m = q\mathbf{v} \times \mathbf{B} \tag{1.8}$$

where $\mathbf{v}$ is the velocity vector of the moving charge, q, and $\mathbf{B} = \mu_0\mathbf{H}$ is the encircling magnetic flux density described by Ørsted. The magnetic field, $\mathbf{H}$, in turn, is given by *Ampère's circuital law*,

$$\boldsymbol{\nabla} \times \mathbf{H} = \mathbf{J} \tag{1.9}$$

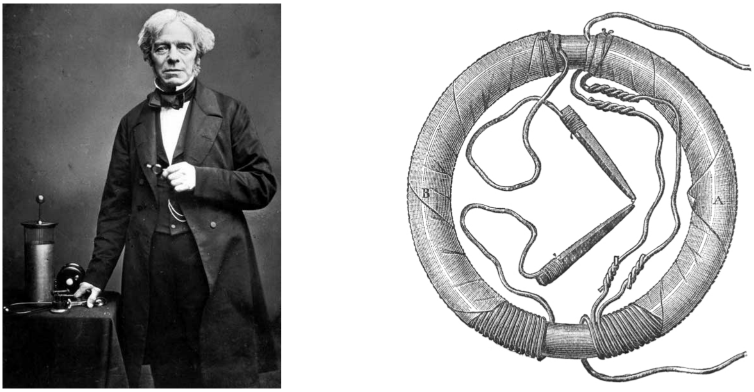
where $\mathbf{J}$ is the current density in space.

Note that we have now identified two forces that act upon charges, either stationary or in motion (Coulomb's and Ampère's, respectively). We may combine these into a single force equation known as the *Lorentz law*,

$$\mathbf{F} = \mathbf{F}_e + \mathbf{F}_m = q\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right) \tag{1.10}$$

1.2.4 Electromagnetic Induction

Ampère's circuital law provides a link between an electric current and the magnetic field that it generates. Another link was found by an English scientist named



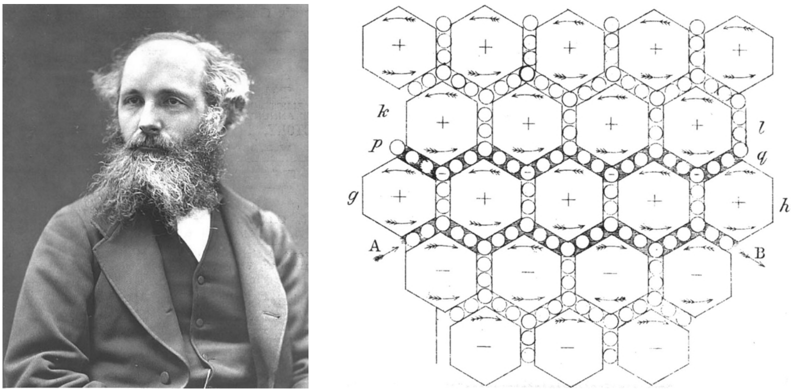
(a) (b)

Figure 1.6 (a) Michael Faraday, English scientist who discovered electromagnetic induction and first proposed the theory of electric and magnetic fields as lines of force. (b) Faraday's ring transformer.

Michael Faraday (Figure 1.6(a)), in 1831. Faraday created what today we would call a toroidal transformer by winding two coils of wire around opposite sides of an iron ring, as in Figure 1.6(b). With the ends of one coil attached to a *galvanometer* (an early type of current detector, or ammeter), he found that connecting the other coil to a chemical battery would induce a momentary current in the first coil, and another momentary current when he disconnected the battery. He further found that swiftly moving a permanent magnet in and out of a loop of wire, or moving that loop through the space around the magnet, would also induce momentary currents in the wire. These are all manifestations of the general theory of electromagnetic induction, for which Faraday is credited [12]. It would eventually be expressed in terms of the electric field and magnetic flux density in the *Maxwell-Faraday equation*, or Faraday's law,

$$\boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{1.11}$$

Faraday's story is an inspiring one. Born to a poor family, he had very little formal education and was impeded at times on account of his low standing in English society. He was nevertheless a brilliant experimentalist whose work profoundly transformed our understanding of the universe — for hidden in Faraday's law of induction is an inescapable truth that Albert Einstein would later cite as a key factor in the development of special relativity [13]. The induced electromotive



(a) (b)

Figure 1.7 (a) James Clerk Maxwell, Scottish scientist who unified the laws of electromagnetics and proposed that light was an electromagnetic wave. (b) Maxwell's model of electromagnetic forces as stresses and strains upon the aether, which was believed to support waves of light.

force in a circuit occurs with the same magnitude whether the magnetic flux through a stationary circuit simply changes, or the circuit itself is physically moved through the field of a static magnetic source. Richard Feynman would later say [14] that "We know of no other place in physics where such a simple and accurate general principle requires for its real understanding an analysis in terms of *two different phenomena*." Despite his humble beginnings, Faraday became one of the most influential scientific minds of all time, and is remembered as a pivotal figure in the development of physics.

It was Faraday who first described electromagnetic phenomena as lines of force extending in space beyond the wires and magnets that produced them. However, this view was not generally accepted by the scientific community in his day, in part because Faraday failed to provide a rigorous mathematical foundation for it. Largely self-taught, his mathematical capabilities were limited to rudimentary algebra, and did not include trigonometry, much less the differential geometry needed to describe vector fields in space.

1.3 MATHEMATICAL REFINEMENT

It was ultimately James Clerk Maxwell, shown in Figure 1.7(a), a Scottish scientist, who provided that mathematical foundation [15, 16]. Maxwell was inquisitive and precocious as a child, and was blessed to be born into a family that had both the means and the will to support his education. He was thus not only the opposite of Faraday in societal standing, but his perfect complement professionally, having a gift for mathematics that was recognized early in life.[5] Maxwell's contribution was not one of experimentation, but of profound theoretical analysis and a remarkable physical analogy [17].

1.3.1 Maxwell's Model

In seeking to unify the vast array of experimental results on electricity and magnetism provided by Faraday and others into a coherent theory, Maxwell drew upon his understanding of hydrodynamics and mechanical systems, fields for which the mathematical tools were already available and were familiar to the scientists of his time [18]. He recognized that the weakening of magnetic forces as the field lines diverged in free space bore a striking mathematical similarity to the manner in which angular momentum decreased as the diameter of a fluid-dynamical vortex grew wider. This and other features convinced him that magnetic forces were essentially rotational in nature.

He therefore conceived of a model in which free space was filled with a sea of corotating vortex tubes that carried the magnetic force. Friction between these tubes was to be mitigated by ball bearings in the form of electric particles occupying the spaces between them (see Figure 1.7(b)). Maxwell openly admitted that this visualization was not intended as a true description of nature, but rather an easily grasped mechanical analogy [19],

> The conception of a particle having its motion connected with that of a vortex by perfect rolling contact may appear somewhat awkward. I do not bring it forward as a mode of connection existing in Nature... It is however a mode of connection which is mechanically conceivable and it serves to bring out the actual mechanical connections between known electromagnetic phenomena.

Nevertheless, Maxwell's model was remarkably successful in reproducing all the known behaviors of electric and magnetic fields in his day. Insulators were characterized as materials in which the electric particles, while free to act as idle wheels for the adjacent magnetic vortices, could not be completely dislocated from

5 He was also known to be somewhat socially awkward, which comes as a comfort to those like myself who are similarly afflicted.

their positions. They could, however, be coaxed into a limited elastic displacement that would account for the storage of electric potential energy in dielectric media. Electromagnetic induction was explained by noting that steady electric currents — manifesting as the uniform flow of electric particles through the channels between vortices — would not cause any movement of particles in the adjacent non-participating channels, unless and until the initial current changed, in which case a transient rotational interaction through the vortices would induce nearby currents in the opposite direction.

Mechanical reasoning of this sort provided the stepping stone for Maxwell to develop and refine his mathematical framework in familiar terms. Ultimately, having served its purpose, the analogy was discarded in favor of the pure abstraction of electromagnetic fields in Maxwell's seminal paper of 1865, "A Dynamical Theory of the Electromagnetic Field" [20]. Like most scientists, Maxwell hoped that by developing a rigorous, structured, and internally consistent theory to describe electromagnetism, it would enable him and others to make verifiable predictions beyond what had already been demonstrated through experimentation. In that, he was wildly successful, for his efforts turned up one of the greatest scientific discoveries of his generation: that light is an electromagnetic wave. To understand how, we must first examine his equations in detail, and, in particular, a term that was missed in all previous experiments, namely, the electric *displacement current*.

1.3.2 Electric Displacement

In modern notation, these are the essential relationships that Maxwell assimilated from others' experiments,

$$\boldsymbol{\nabla} \cdot \mathbf{D} = \rho \tag{1.12a}$$

$$\boldsymbol{\nabla} \cdot \mathbf{B} = 0 \tag{1.12b}$$

$$\boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{1.12c}$$

$$\boldsymbol{\nabla} \times \mathbf{H} = \mathbf{J} \tag{1.12d}$$

The first is Gauss's law, seen earlier in (1.6c), and it is substantiated by Coulomb's measurements of the electrostatic force. The second equation, originally (1.1), expresses the nonexistence of magnetic monopoles, one of the fundamental principles of magnets espoused by William Gilbert. Third is Faraday's law of induction, also shown in (1.11).

The last equation is Ampère's circuital law, (1.9), which, indirectly through the action of the magnetic field, describes the force exerted by a current-carrying

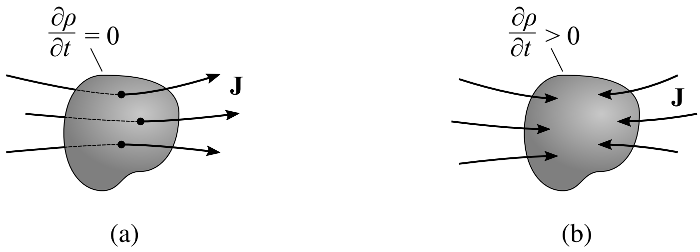


Figure 1.8 Continuity of electric charge density (a) without the displacement current, and (b) with the displacement current.

wire upon another that is parallel to it — but this law is incomplete. The issue is revealed if we take the divergence of both sides, and note that the divergence of a curl is always zero,

$$\nabla \cdot \mathbf{J} = \nabla \cdot \nabla \times \mathbf{H} = 0 \tag{1.13}$$

Ampère's law as it is given would predict, incorrectly, that the current density, $\mathbf{J}$, has no divergence, or, equivalently, that charges cannot accumulate. Draw a boundary around any arbitrary region of space, as in Figure 1.8(a), and the current entering it is exactly the same as the current leaving; no more charge can go in than comes out. This is obviously incorrect, as the accumulation of charge is the very basis of the triboelectric effect, one of the first evidences of electric phenomena that has been observed for thousands of years, not to mention the measurements of Coulomb and others who calculated the force exerted by charges accumulated on the electrodes of their apparatuses.

To repair this, Maxwell found it necessary to add a term known as the displacement current [21–23], $\partial\mathbf{D}/\partial t$, to Ampère's law,

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{1.14}$$

Then, when we take the divergence, and substitute for the displacement field, $\mathbf{D}$, from Gauss's law,

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t} \tag{1.15}$$

We now predict that a diverging current density in a region of space must be accompanied by an ongoing depletion of charge from that region, and conversely, if the divergence is negative (the current density is convergent), then there must be charge accumulating in that region, as illustrated in Figure 1.8(b). This is exactly what you would expect if charge is a conserved quantity; charge cannot be created nor destroyed, but only moved from place to place. Numerous physical experiments had already confirmed this. Equation (1.15) is known as the *continuity equation*.

However, the implications of Maxwell's displacement current go far beyond ensuring the continuity of charge. Like all the greatest scientific achievements, it answered a profound fundamental question, "What is the nature of light?" while at the same time posing another even more profound question, "How can light's velocity be frame-independent?" We will answer the first question in the following sections. Consideration of the second question will be taken up in Chapter 2.

1.3.3 Propagating Waves

Consider what happens if we take the curl of both sides of Faraday's law,

$$\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial}{\partial t}(\boldsymbol{\nabla} \times \mathbf{B}) \tag{1.16}$$

Then, apply the vector identity (B.35) and substitute from Ampère's law,

$$\boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{E}) - \boldsymbol{\nabla}^2 \mathbf{E} = -\mu_0 \frac{\partial}{\partial t}\left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\right) \tag{1.17}$$

Finally, if we assume that the region is source-free ($\rho = 0$ and $\mathbf{J} = 0$), we are left with

$$\boldsymbol{\nabla}^2 \mathbf{E} = \mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{1.18}$$

Example: This is a form of the wave equation in three dimensions. It has numerous possible solutions depending on the geometry of the initial conditions, but we may articulate one very simple case by assuming that $\mathbf{E}$ is uniform in the yz plane (changing only in x and t) and directed along the z axis,

$$\mathbf{E} = E(x, t)\mathbf{z} \tag{1.19a}$$

$$\therefore\ \boldsymbol{\nabla}^2 \mathbf{E} = \frac{\partial^2 E}{\partial x^2}\mathbf{z} = \mu_0 \varepsilon_0 \frac{\partial^2 E}{\partial t^2}\mathbf{z} \tag{1.19b}$$

which has solutions of the form,

$$\mathbf{E} = E(x \pm ct)\mathbf{z} \tag{1.20}$$

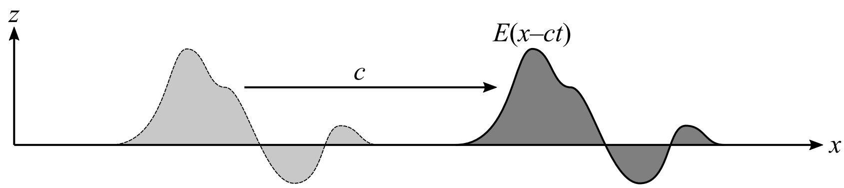


Figure 1.9 A rightward-traveling wave.

where

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} \tag{1.21}$$

In other words, the electric field may take on any smooth (differentiable) functional form along the x-axis, but that waveform will then propagate left or right at speed c, known as the *phase velocity*. A rightward-traveling wave is thus illustrated in Figure 1.9. Solutions of this type are called *plane waves*. Imagine Maxwell's delight upon substituting the values for the permittivity and permeability of empty space which had been measured experimentally by other researchers, and obtaining the following result [18],

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} \approx 310,740 \text{ km/s} \tag{1.22}$$

In Maxwell's own words [24],

> The velocity of transverse modulations in our hypothetical medium, calculated from the electro-magnetic experiments of MM. Kohlrausch and Weber, agrees so exactly with the velocity of light calculated from the optical experiments of M. Fizeau that we can scarcely avoid the inference that *light consists in the transverse modulations of the same medium which is the cause of electric and magnetic phenomena.*

To Maxwell, the concurrence of the theoretically calculated propagation speed of electromagnetic waves with the experimentally measured velocity of light to a precision of about 1%, well within experimental error,[6] could not be dismissed as a coincidence; light is an electromagnetic wave. This unification of the previously

6 Fizeau had determined the speed of light to be 313,300 km/s [25], within 1% of Maxwell's calculation. Today, more precise measurements along with refined SI units purposefully designed to eliminate the approximation yield an exact value of $c = 299,792,458$ m/s, giving Fizeau's and Maxwell's figures an error of about 4.5% and 3.5%, respectively.

distinct phenomena of light and electromagnetism is now recognized as one of the greatest triumphs of mathematical physics.

1.3.4 Potential Formulation

The form in which the electric and magnetic fields appear in Maxwell's equations makes it possible to express them as derivatives of potential functions, a concept that Maxwell took advantage of in his publications.[7] Because the magnetic flux density is divergence-free, for example, we may write it as the curl of a magnetic vector potential, $\mathbf{A}$,

$$\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A} \tag{1.23}$$

Since the divergence of a curl is always zero, (1.1) is automatically satisfied. Similarly, in the static case where time derivatives are zero, Faraday's law tells us that the electric field has no curl,

$$\boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} = 0 \tag{1.24}$$

We may thus write the electrostatic field as the *gradient* of a scalar potential, φ, for the curl of all gradients is also guaranteed to be zero,

$$\mathbf{E} = -\boldsymbol{\nabla}\varphi \tag{1.25}$$

The scalar field φ is known as the voltage, and the minus sign is a convention that ensures that the electric field vector always points from high to low potentials. In the nonstatic case, where the time derivatives are not zero, the electric field is not entirely curl-free; there is an additional component given by

$$-\frac{\partial \mathbf{B}}{\partial t} = -\frac{\partial}{\partial t}(\boldsymbol{\nabla} \times \mathbf{A}) \tag{1.26}$$

Therefore, to be consistent with Maxwell's equations in the most general case, we must write the full electric field as

$$\mathbf{E} = -\boldsymbol{\nabla}\varphi - \frac{\partial \mathbf{A}}{\partial t} \tag{1.27}$$

7 Maxwell's enthusiasm for the field potentials was not shared by Oliver Heaviside, who reduced Maxwell's twenty equations into four using the vector operator notation still in use today. To Heaviside, the electrostatic potential was a "physical inanity," signifying nothing real, and declared that it was "best to murder the whole lot" [26].

Note that this reduces to the simple gradient form in the static case.

With the definitions given in (1.23) and (1.27), both Faraday's law and the nonexistence of magnetic charge are guaranteed to be satisfied by mathematical form. Substitution into Ampère's law (including Maxwell's displacement current) gives

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{1.28a}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \tag{1.28b}$$

$$\nabla \times \nabla \times \mathbf{A} = \mu_0 \mathbf{J} - \frac{1}{c^2} \frac{\partial}{\partial t} \left(\nabla \varphi + \frac{\partial \mathbf{A}}{\partial t} \right) \tag{1.28c}$$

Let us once again apply the vector identity (B.35),

$$\nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = \mu_0 \mathbf{J} - \frac{1}{c^2} \frac{\partial}{\partial t} \left(\nabla \varphi + \frac{\partial \mathbf{A}}{\partial t} \right) \tag{1.29a}$$

$$\left(\frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla^2 \mathbf{A} \right) + \nabla \left(\nabla \cdot \mathbf{A} + \frac{1}{c^2} \frac{\partial \varphi}{\partial t} \right) = \mu_0 \mathbf{J} \tag{1.29b}$$

This is Ampère's law in potential form [21]. Substitution into Gauss's law gives

$$\nabla \cdot \mathbf{D} = \varepsilon_0 \nabla \cdot \mathbf{E} = \varepsilon_0 \left(-\nabla^2 \varphi - \frac{\partial}{\partial t} (\nabla \cdot \mathbf{A}) \right) = \rho \tag{1.30a}$$

$$\therefore \nabla^2 \varphi + \frac{\partial}{\partial t} (\nabla \cdot \mathbf{A}) = -\frac{\rho}{\varepsilon_0} \tag{1.30b}$$

By using potentials, we have thus reduced the number of Maxwell's equations from four to two — that is, these two equations together with the defining relationships between the potentials and electromagnetic fields, (1.23) and (1.27), guarantee that the original four of what today we call Maxwell's equations are satisfied. The potential forms are more complex than the originals, Ampère's law especially, but there is still some room to make simplifications. Importantly, the scalar and vector potentials are not entirely determined by (1.29b) and (1.30b); more than one set of functions, φ and $\mathbf{A}$, will satisfy them, and thus indirectly determine the unique values of $\mathbf{E}$ and $\mathbf{H}$. One could, for example, add a constant offset to the scalar potential, φ, which appears only in differentiated terms in these equations. Since only $\mathbf{E}$ and $\mathbf{H}$ can be directly measured, the differences between any such valid choices of φ and $\mathbf{A}$ have no physical implications. This ambiguity is known as *gauge freedom*, and the selection of an additional criterion to resolve that ambiguity is known as *gauge fixing*.

Various gauge conditions have been used in different situations. The *Coulomb gauge* makes the assumption that $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$; thus, we obtain

$$\frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} - \boldsymbol{\nabla}^2 \mathbf{A} = \mu_0 \mathbf{J} - \frac{1}{c^2}\boldsymbol{\nabla}\frac{\partial \varphi}{\partial t} \tag{1.31a}$$

$$\boldsymbol{\nabla}^2 \varphi = -\frac{\rho}{\varepsilon_0} \tag{1.31b}$$

For our purposes, the *Lorenz gauge*[8] is more appropriate,

$$\boldsymbol{\nabla} \cdot \mathbf{A} = -\frac{1}{c^2}\frac{\partial \varphi}{\partial t} \tag{1.32}$$

in which case, Maxwell's equations become

$$\frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} - \boldsymbol{\nabla}^2 \mathbf{A} = \mu_0 \mathbf{J} \tag{1.33a}$$

$$\frac{1}{c^2}\frac{\partial^2 \varphi}{\partial t^2} - \boldsymbol{\nabla}^2 \varphi = \frac{\rho}{\varepsilon_0} \tag{1.33b}$$

or, in terms of the *d'Alembertian* operator,[9]

$$\Box^2 = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \boldsymbol{\nabla}^2 \tag{1.34a}$$

$$= \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2} \tag{1.34b}$$

we may write

$$\Box^2 \mathbf{A} = \mu_0 \mathbf{J} \tag{1.35a}$$

$$\Box^2 \varphi = \frac{\rho}{\varepsilon_0} \tag{1.35b}$$

These equations look quite similar to one another, and even more importantly, the two potential fields have become decoupled. Both are forms of the wave equation

8 Those readers who think Lorenz is misspelled here might be interested in reading Appendix F.

9 Some texts write the d'Alembertian operator as a single box, $\Box$, instead of the box-squared, $\Box^2$. We shall reserve the single box notation for another operator, called the *spacetime gradient*, whose square is indeed the d'Alembertian given here.

where the current density, $\mathbf{J}$, appears as the source of waves in the magnetic vector potential, and the charge density, ρ, appears as the source of waves in the scalar electric potential, both propagating at the speed of light, c. This decoupling is precisely the value of using the potentials. Even using the Coulomb gauge, one could solve (1.31b) for φ first, then (1.31a) for $\mathbf{A}$, then plug those results directly into (1.27) and (1.23) independently to determine $\mathbf{E}$ and $\mathbf{H}$. The equations are solved sequentially, whereas the four canonical Maxwell's equations are coupled in such a way that they usually have to be solved simultaneously, which in many cases is far more difficult.

Note that in the earlier form derived from the Coulomb gauge, (1.31), the scalar electric potential responds instantly at all points throughout space, no matter how far removed from the source charge density, whereas the magnetic vector potential propagates with a finite speed, c. Nevertheless, since the electric field depends on both potentials (and the magnetic field on the vector potential alone), these fields still have a propagation delay limited by the speed of light.

It appears that no matter how we choose to formulate Maxwell's equations, there is no way for a localized disturbance of charges or currents to affect the electromagnetic fields (quantities which, unlike the potentials, are directly measurable) at a distance any more rapidly than the travel time of light between them.

A summary of the various differential forms of Maxwell's equations described in this chapter appears in Table 1.1.

1.4 MATTER AND ENERGY

Equipped now with the mathematical tools to describe the fundamental electromagnetic forces and the behavior of elemental particles in response to those forces, we may begin to apply them to concepts of a slightly more practical nature, namely the interaction of electromagnetic fields with matter on a macroscopic scale, and the ebb and flow of energy in electromagnetic systems.

1.4.1 Material Constituent Parameters

Thus far, we have confined our discussion to the behavior of electromagnetic fields in the vacuum of space (and implicitly assumed that the behavior of these fields in Earth's atmosphere would not be much different). In a strict sense, this is perfectly valid, since even insulating materials comprise discrete molecules that are separated from one another by empty space. In principle, the electromagnetic

Table 1.1
Maxwell's Equations in Free Space

Name	Expression
Direct form:	
Gauss's law	$\boldsymbol{\nabla} \cdot \mathbf{D} = \rho$
Magnetic nondivergence	$\boldsymbol{\nabla} \cdot \mathbf{B} = 0$
Faraday's law	$\boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$
Ampère's law	$\boldsymbol{\nabla} \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}$
Potential form, gauge-free:	
Ampère's law	$\left(\frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} - \boldsymbol{\nabla}^2 \mathbf{A}\right) + \boldsymbol{\nabla}\left(\boldsymbol{\nabla} \cdot \mathbf{A} + \frac{1}{c^2}\frac{\partial \varphi}{\partial t}\right) = \mu_0 \mathbf{J}$
Gauss's law	$\boldsymbol{\nabla}^2 \varphi + \frac{\partial}{\partial t}(\boldsymbol{\nabla} \cdot \mathbf{A}) = -\frac{\rho}{\varepsilon_0}$
Potential form, Coulomb gauge:	
Gauge condition	$\boldsymbol{\nabla} \cdot \mathbf{A} = 0$
Ampère's law	$\frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} - \boldsymbol{\nabla}^2 \mathbf{A} = \mu_0 \mathbf{J} - \frac{1}{c^2}\boldsymbol{\nabla}\frac{\partial \varphi}{\partial t}$
Gauss's law	$\boldsymbol{\nabla}^2 \varphi = -\frac{\rho}{\varepsilon_0}$
Potential form, Lorenz gauge:	
Gauge condition	$\boldsymbol{\nabla} \cdot \mathbf{A} = -\frac{1}{c^2}\frac{\partial \varphi}{\partial t}$
Ampère's law	$\Box^2 \mathbf{A} = \mu_0 \mathbf{J}$
Gauss's law	$\Box^2 \varphi = \frac{\rho}{\varepsilon_0}$

fields in these interstitial spaces could be calculated from the combined effect of the charged protons and electrons within those molecules responding to external stimuli. However, for many practical situations, this approach would be hopelessly complex, or, at best, needlessly tedious. For that reason, engineers have devised a shortcut.

Consider the simple model of a dielectric material in Figure 1.10(a). The molecules are drawn as spheres having positive charge in the middle (shaded black) surrounded by a cloud of negative charge (shaded white) such that each is electrically neutral and symmetric. The material is said to be unbiased, or unpolarized. When an external electric field is applied, as in Figure 1.10(b), these positive and negative charges are pulled away from each other. Although they are not fully dissociated from one another, the partial separation creates a tiny electric dipole whose field lines point in the direction opposite to the applied field. In effect, the material partially repels the electric field, reducing the force that a charged

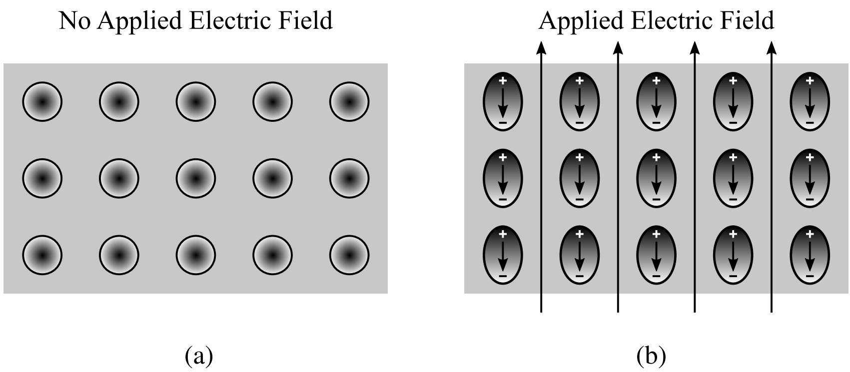


Figure 1.10 Illustration of dielectric polarizability. (a) Unbiased material comprising electrically neutral molecules. (b) The application of an external bias field polarizes those molecules, drawing the positive and negative charges apart, and setting up electric dipole moments which tend to counteract the applied field in their vicinity.

test particle that is free to move through the material would feel. The extra flux density, $\Delta\mathbf{D}$, that is required to produce a given net electric force, $\mathbf{E}$, is known as the *dielectric polarization density*, and may be written

$$\mathbf{P} = \Delta\mathbf{D} = \varepsilon_0 \chi_e \mathbf{E} \tag{1.36}$$

where χ_e is a dimensionless material parameter called the *electric susceptibility*. We may therefore write the total electric flux density as follows,

$$\mathbf{D} = \varepsilon_0 \left(1 + \chi_e\right) \mathbf{E} = \varepsilon_0 \varepsilon_r \mathbf{E} = \varepsilon \mathbf{E} \tag{1.37}$$

The scaling factor, ε_r, is called the *relative permittivity* or *dielectric constant*. We may thus rewrite all of Maxwell's equations using a net effective permittivity, ε, where the only source charge densities that are included are those that are free (not bound to the material lattice), sometimes denoted ρ_f. Bound charges are already accounted for in the effective permittivity.

A similar analysis shows that magnetic dipole moments within certain materials align with an externally applied magnetic field, such that the magnetic flux density is also increased. In this case, the magnetization, $\mathbf{M}$, is given by

$$\mathbf{M} = \chi_m \mathbf{H} \tag{1.38}$$

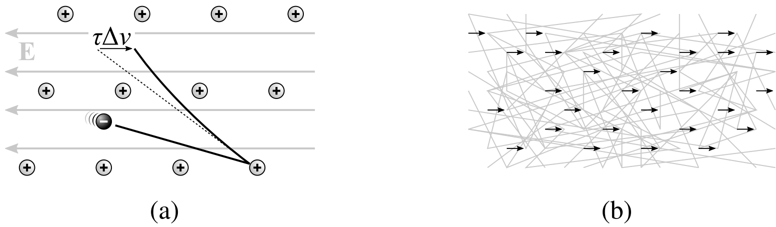


Figure 1.11 Illustration of the mechanism responsible for conductivity in metals. (a) Small perturbation of a single electron's velocity vector between events. (b) Cumulative effect of velocity perturbations averaged over many charge carriers and scattering events.

where χ_m is the *magnetic susceptibility*. The net effective magnetic flux density is then

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) = \mu_0 (1 + \chi_m) \mathbf{H} = \mu_0 \mu_r \mathbf{H} = \mu \mathbf{H} \tag{1.39}$$

The parameter μ_r is the *relative permeability*, and μ is the net effective permeability of the material, which may be substituted into Maxwell's equations to encapsulate the effect of the bound currents within the material, leaving only the free currents, $\mathbf{J}_f$, to be dealt with explicitly.

It must be emphasized that Maxwell's equations, when written with these parameters — that is, with ε and μ instead of ε_0 and μ_0 — describe average macroscopic effects only, while smoothing out the microscopic details. This is more than sufficient for most engineering problems, but it is, in reality, only an approximation of the truth. We must therefore be very careful about what conclusions we draw from the above analyses in a fundamental sense. We will discuss the implications of this in more detail in Section 2.2.1, but for now, let us play it safe and work only with the microscopically accurate form of Maxwell's equations, using the constituent parameters of the vacuum, ε_0 and μ_0.

1.4.2 Conductivity

The origin of conductivity in metals is also an average effect brought about by the behavior of many small charge carriers acting in concert. Picture a single electron moving through a material lattice of positive atomic nuclei, as shown in Figure 1.11(a). Even in the absence of an externally applied electric field, charge carriers such as this electron will already be in motion at random as a consequence of thermal excitation. Periodically, the electron's path will be randomized by scattering, or collisions with the lattice nuclei. Averaged over many cycles and over a large number of carriers all moving at once, the net effective current in the absence of any ambient electric field will be zero (however, there will be minute fluctuations of current around zero due to the statistical nature of this phenomenon, which manifests in real-world systems as electronic noise).

If there is an electric field present, it will tend to accelerate the charge carriers in the interim period between collisions, bending their paths in a preferred direction. Note that the electron's path in Figure 1.11(a) has deviated to the right as a consequence of the leftward-directed electric field and the particle's negative charge. Under typical conditions (such as temperature), the time period for acceleration between scattering events is so short that the change in velocity imparted to the particle by Lorentz forces is less than its random thermal motion by about 10 orders of magnitude [27]. Thus, the mean free time between collisions is overwhelmingly dominated by temperature, and is virtually unaffected by the application of an electric field.

Nevertheless, these tiny changes in velocity, averaged over a large number of particles scattering over and over again, add coherently into a net effective drift of charges in a common direction, as shown in Figure 1.11(b). The rate at which the cloud of charges drifts is proportional to the field strength multiplied by the mean free time between collisions (which, as stated above, is virtually independent of field strength). Therefore, the net effective current is essentially proportional to the electric field. This relationship, which is more empirical than absolute, is known as *Ohm's law*,

$$\mathbf{J} = \sigma \mathbf{E} \tag{1.40}$$

where the proportionality constant, σ, is the conductivity of the material.

1.4.3 Conservation of Energy

As we know, an electric field is created by the accumulation or separation of electric charges, but those charges resist being accumulated or separated as a consequence of the Lorentz force they exert upon one another. It therefore requires energy to produce an electric field. It can be shown [21] that the energy thus stored in a static electric field per unit volume is

$$u_E = \tfrac{1}{2}\mathbf{E} \cdot \mathbf{D} \tag{1.41}$$

Similarly, a magnetic field is created by the motion of electric charges, and energy is required to set those charges in motion. The energy density stored in a magnetic field is

$$u_H = \tfrac{1}{2}\mathbf{B} \cdot \mathbf{H} \tag{1.42}$$

Let us now consider the rate at which the total energy density (electric and magnetic) changes,

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial t}(u_E + u_H) = \frac{1}{2}\frac{\partial}{\partial t}(\mathbf{E}\cdot\mathbf{D} + \mathbf{B}\cdot\mathbf{H}) \tag{1.43a}$$

$$= \frac{1}{2}\left(\mathbf{E}\cdot\frac{\partial \mathbf{D}}{\partial t} + \mathbf{D}\cdot\frac{\partial \mathbf{E}}{\partial t} + \mathbf{H}\cdot\frac{\partial \mathbf{B}}{\partial t} + \mathbf{B}\cdot\frac{\partial \mathbf{H}}{\partial t}\right) = \mathbf{E}\cdot\frac{\partial \mathbf{D}}{\partial t} + \mathbf{H}\cdot\frac{\partial \mathbf{B}}{\partial t} \tag{1.43b}$$

where we have allowed the underlying constants μ_0 and ε_0 to move in and out of the derivatives so that we may combine terms. We may solve the partial time derivatives above using Faraday's and Ampère's laws,

$$\frac{\partial u}{\partial t} = \mathbf{E}\cdot(\boldsymbol{\nabla}\times\mathbf{H} - \mathbf{J}) + \mathbf{H}\cdot(-\boldsymbol{\nabla}\times\mathbf{E}) \tag{1.44a}$$

$$= \mathbf{E}\cdot(\boldsymbol{\nabla}\times\mathbf{H}) - \mathbf{H}\cdot(\boldsymbol{\nabla}\times\mathbf{E}) - \mathbf{J}\cdot\mathbf{E} \tag{1.44b}$$

$$= -\boldsymbol{\nabla}\cdot(\mathbf{E}\times\mathbf{H}) - \mathbf{J}\cdot\mathbf{E} \tag{1.44c}$$

where the last step follows from the vector identity (B.33). The last term, $\mathbf{J}\cdot\mathbf{E}$, represents the work done by the Lorentz force acting upon charges having inertial mass. The remaining, divergent term evidently represents power flow in the propagation of the electromagnetic fields themselves, absent the motion of massive charges. The *cross product* of the electric and magnetic fields, $\mathbf{S} = \mathbf{E}\times\mathbf{H}$, is termed the *Poynting vector*. The resulting equation, known as *Poynting's theorem*,

$$-\frac{\partial u}{\partial t} = \boldsymbol{\nabla}\cdot\mathbf{S} + \mathbf{J}\cdot\mathbf{E} \tag{1.45}$$

is thus effectively an expression of the conservation of electromagnetic energy. It balances the depletion of stored energy ($-\partial u/\partial t$) in a given region of space on the left, with power flowing out ($\boldsymbol{\nabla}\cdot\mathbf{S}$) and dissipated ($\mathbf{J}\cdot\mathbf{E}$) in that region on the right.

1.4.4 Conservation of Momentum

In Newtonian mechanics, force is mass times acceleration, or, more succinctly, it is the rate of change of momentum. In a closed system, momentum should be conserved. Let us examine the total force of an electromagnetic system through the Lorentz force law,

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v}\times\mathbf{B}) = q\mathbf{E} + (q\mathbf{v})\times\mathbf{B} \tag{1.46}$$

In the above equation, q represents a discrete lump of electric charge. To be more general, let us divide through by volume to turn this into a charge density, $q \to \rho$. The product of that charge density with velocity is then a current density, $q\mathbf{v} \to \mathbf{J}$. The force per unit volume, $\mathbf{f}$, is then

$$\mathbf{f} = \rho\mathbf{E} + \mathbf{J} \times \mathbf{B} \tag{1.47}$$

We may eliminate the source terms by substituting from Gauss's and Ampère's laws,

$$\mathbf{f} = (\boldsymbol{\nabla} \cdot \mathbf{D})\mathbf{E} + \left(\boldsymbol{\nabla} \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t}\right) \times \mathbf{B} \tag{1.48a}$$

$$= (\boldsymbol{\nabla} \cdot \mathbf{D})\mathbf{E} + (\boldsymbol{\nabla} \times \mathbf{H}) \times \mathbf{B} - \frac{\partial \mathbf{D}}{\partial t} \times \mathbf{B} \tag{1.48b}$$

The final term involving a time derivative may be associated with the Poynting vector as follows [28],

$$\frac{\partial \mathbf{S}}{\partial t} = \frac{\partial}{\partial t}(\mathbf{E} \times \mathbf{H}) = \frac{\partial \mathbf{E}}{\partial t} \times \mathbf{H} + \mathbf{E} \times \frac{\partial \mathbf{H}}{\partial t} = \frac{\partial \mathbf{E}}{\partial t} \times \mathbf{H} - \frac{1}{\mu_0}\mathbf{E} \times (\boldsymbol{\nabla} \times \mathbf{E}) \tag{1.49a}$$

$$\therefore \frac{\partial \mathbf{D}}{\partial t} \times \mathbf{B} = \varepsilon_0\mu_0 \left(\frac{\partial \mathbf{S}}{\partial t} + \frac{1}{\mu_0}\mathbf{E} \times (\boldsymbol{\nabla} \times \mathbf{E})\right) \tag{1.49b}$$

where we have also applied Faraday's law. Therefore, returning to the force density,

$$\mathbf{f} = (\boldsymbol{\nabla} \cdot \mathbf{D})\mathbf{E} + (\boldsymbol{\nabla} \times \mathbf{H}) \times \mathbf{B} - \varepsilon_0\mu_0 \left(\frac{\partial \mathbf{S}}{\partial t} + \frac{1}{\mu_0}\mathbf{E} \times (\boldsymbol{\nabla} \times \mathbf{E})\right) \tag{1.50a}$$

$$= (\boldsymbol{\nabla} \cdot \mathbf{D})\mathbf{E} - \mathbf{D} \times (\boldsymbol{\nabla} \times \mathbf{E}) - \mathbf{B} \times (\boldsymbol{\nabla} \times \mathbf{H}) - \frac{1}{c^2}\frac{\partial \mathbf{S}}{\partial t} \tag{1.50b}$$

The electric and magnetic fields would appear symmetrically in the above equation if there were a term $(\boldsymbol{\nabla} \cdot \mathbf{B})\mathbf{H}$. The nonexistence of magnetic charge guarantees that $\boldsymbol{\nabla} \cdot \mathbf{B} = 0$, so adding it does not change anything,

$$\mathbf{f} = (\boldsymbol{\nabla} \cdot \mathbf{D})\mathbf{E} - \mathbf{D} \times (\boldsymbol{\nabla} \times \mathbf{E}) + (\boldsymbol{\nabla} \cdot \mathbf{B})\mathbf{H} - \mathbf{B} \times (\boldsymbol{\nabla} \times \mathbf{H}) - \frac{1}{c^2}\frac{\partial \mathbf{S}}{\partial t} \tag{1.51}$$

The curls in the above equation can be eliminated using the vector identity (B.32), which expressed in a simpler form is

$$\mathbf{A} \times (\boldsymbol{\nabla} \times \mathbf{A}) = \tfrac{1}{2}\boldsymbol{\nabla}(\mathbf{A} \cdot \mathbf{A}) - (\mathbf{A} \cdot \boldsymbol{\nabla})\mathbf{A} \tag{1.52}$$

Thus,

$$\mathbf{f} = (\boldsymbol{\nabla} \cdot \mathbf{D})\mathbf{E} + (\mathbf{E} \cdot \boldsymbol{\nabla})\mathbf{D} + (\boldsymbol{\nabla} \cdot \mathbf{B})\mathbf{H} + (\mathbf{H} \cdot \boldsymbol{\nabla})\mathbf{B} - \boldsymbol{\nabla} u - \frac{1}{c^2}\frac{\partial \mathbf{S}}{\partial t} \tag{1.53}$$

where u is the electromagnetic energy density as described in Section 1.4.3.[10] This can be written most compactly if we define the *Maxwell stress tensor*, a 3×3 matrix,

$$\boldsymbol{\sigma} = \mathbf{E} \otimes \mathbf{D} + \mathbf{B} \otimes \mathbf{H} - u\mathbf{I} \tag{1.54a}$$

$$= \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} \begin{pmatrix} D_x & D_y & D_z \end{pmatrix} + \begin{pmatrix} B_x \\ B_y \\ B_z \end{pmatrix} \begin{pmatrix} H_x & H_y & H_z \end{pmatrix} - u\mathbf{I} \tag{1.54b}$$

$$= \begin{pmatrix} E_x D_x + B_x H_x & E_x D_y + B_x H_y & E_x D_z + B_x H_z \\ E_y D_x + B_y H_x & E_y D_y + B_y H_y & E_y D_z + B_y H_z \\ E_z D_x + B_z H_x & E_z D_y + B_z H_y & E_z D_z + B_z H_z \end{pmatrix} - u\mathbf{I} \tag{1.54c}$$

where $\otimes$ is called the *dyadic product* (see Section B.1.4) and $\mathbf{I}$ is the 3×3 identity matrix. One may observe that the tensor divergence of the Maxwell stress tensor accounts for all terms of the force density except for that involving the Poynting vector. For example, if we distribute the del operator across both terms of the *outer product* $\mathbf{E} \otimes \mathbf{D}$ using the product rule for differentiation (being careful to preserve the product order, since dyadic products are noncommutative),

$$\boldsymbol{\nabla} \cdot (\mathbf{E} \otimes \mathbf{D}) = (\boldsymbol{\nabla} \cdot \mathbf{E}) \otimes \mathbf{D} + \mathbf{E} \cdot (\boldsymbol{\nabla} \otimes \mathbf{D}) = (\boldsymbol{\nabla} \cdot \mathbf{E})\mathbf{D} + (\mathbf{E} \cdot \boldsymbol{\nabla})\mathbf{D} \tag{1.55}$$

and similarly for $\boldsymbol{\nabla} \cdot (\mathbf{B} \otimes \mathbf{H})$. Thus, we have

$$\boldsymbol{\nabla} \cdot \boldsymbol{\sigma} = \mathbf{f} + \frac{1}{c^2}\frac{\partial \mathbf{S}}{\partial t} \tag{1.56}$$

Although we will not see this particular equation again, its physical interpretation is of the utmost importance. It is a law of the conservation of momentum coupling both mechanical and electromagnetic forces. As stated at the beginning of this section, the force, $\mathbf{f}$, is the rate of change of momentum (or rather, in this case, the momentum density) for inertially massive particles. The term involving the derivative of the Poynting vector may then be interpreted as a momentum for

10 It should be noted that some of the steps in this derivation depended on moving the constant constituent parameters of the vacuum, ε_0 and μ_0, in and out of spatial derivatives. The same would apply in the macroscopic case using ε and μ so long as the material was isotropic.

the electromagnetic field itself. In other words, (1.56) affirms that electromagnetic waves carry momentum, a striking conclusion since electromagnetic waves have no mass (in our modern understanding, we would say that photons have no *rest mass*). Light, as an electromagnetic wave, exerts a small but measurable pressure upon the surfaces on which it is incident, foreshadowing the conclusions of special relativity, as we shall see later in this book.

The Maxwell stress tensor frequently reoccurs elsewhere in electromagnetic analysis. I shall have more to say about tensors in general in Chapter 4.

References

[1] Wikipedia. (2024, January) Triboelectric effect. https://en.wikipedia.org/wiki/Triboelectric_effect.

[2] Wikipedia. (2024, January) Thales of Miletus. https://en.wikipedia.org/wiki/Thales_of_Miletus.

[3] Wikipedia. (2023, April) Lodestone. https://en.wikipedia.org/wiki/Lodestone.

[4] S. Winchester, *The Perfectionists: How Precision Engineers Created the Modern World.* New York: Harper Collins, 2018.

[5] J. Bryant and C. Sangwin, *How Round is Your Circle?: Where Engineering and Mathematics Meet*. New Jersey: Princeton University Press, 2008.

[6] Wikipedia. (2024, January) William Gilbert (astronomer). https://en.wikipedia.org/wiki/William_Gilbert_(astronomer).

[7] G. Arfken and H. Weber, *Mathematical Methods for Physicists*, 4th ed. San Diego, CA: Academic Press, 1995.

[8] Wikipedia. (2024, January) Coulomb's law. https://en.wikipedia.org/wiki/Coulomb%27s_law.

[9] R. de Andrade Martins, "Resistance to the discovery of electromagnetism: Ørsted and the symmetry of the magnetic field," *Volta and the History of Electricity*, pp. 245–265, 2003.

[10] Wikipedia. (2023, November) Hans Christian Ørsted. https://en.wikipedia.org/wiki/Hans_Christian_%C3%98rsted.

[11] Wikipedia. (2024, January) André-Marie Ampère. https://en.wikipedia.org/wiki/Andr%C3%A9-Marie_Amp%C3%A8re.

[12] Wikipedia. (2023, December) Faraday's law of induction. https://en.wikipedia.org/wiki/Faraday%27s_law_of_induction.

[13] A. Einstein, "Zur elektrodynamik bewegter körper (on the electrodynamics of moving bodies)," *Annalen der Physik*, vol. 322, no. 10, pp. 891–921, June 1905.

[14] R. P. Feynman, *The Feynman Lectures on Physics*, R. B. Leighton and M. Sands, Eds. San Francisco: Pearson/Addison-Wesley, 2006.

[15] Wikipedia. (2024, January) James Clerk Maxwell. https://en.wikipedia.org/wiki/James_Clerk_Maxwell.

[16] R. P. Crease, *The Great Equations*. New York: Norton, 2010.

[17] I. Stewart, *In Pursuit of the Unknown: 17 Equations That Changed the World*. New York: Basic Books, 2012.

[18] M. Longair, "'…a paper…I hold to be great guns': a commentary on Maxwell (1865) 'A dynamical theory of the electromagnetic field'," *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences*, vol. 373, no. 2039, pp. 1–22, April 2015.

[19] J. C. Maxwell, "On physical lines of force. II. The theory of molecular vortices applied to electric currents," *Philosophical Magazine*, vol. 21, pp. 338–348, 1861.

[20] J. C. Maxwell, "A dynamical theory of the electromagnetic field," *Philosophical Transactions of the Royal Society of London*, vol. 155, pp. 459–512, 1865.

[21] M. A. Morgan, *Principles of RF and Microwave Design*. Norwood, MA: Artech House, November 2019.

[22] S. Ramo, J. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics*. New York: Wiley, 1984.

[23] D. M. Pozar, *Microwave Engineering*, 4th ed. New York: Wiley, 2011.

[24] J. C. Maxwell, "On physical lines of force. III. The theory of molecular vortices applied to electric currents," *Philosophical Magazine*, vol. 23, pp. 12–24, 1861.

[25] Wikipedia. (2023, August) Hippolyte Fizeau. https://en.wikipedia.org/wiki/Hippolyte_Fizeau.

[26] P. J. Nahin, *Oliver Heaviside: The Life, Works, and Times of an Electrical Genius of the Victorian Age*. Baltimore, MD: Johns Hopkins University Press, 2002.

[27] M. Schwartz, *Principles of Electrodynamics*. New York: Dover, 1987.

[28] Wikipedia. (2019, September) Maxwell stress tensor. https://en.wikipedia.org/wiki/Maxwell_stress_tensor.

Chapter 2

Reference Frame Transformation

The principle of relativity[1] states that physical laws must be the same to observers in any permissible frame of reference. That is not to say that different observers do not have different perspectives as to the specific state of objects in a given system — a cup of coffee on a train may appear to be moving rapidly sideways to those outside viewing it through a window, but not so to the passenger inside who sets it on the table before him with confidence that it will not turn over — rather, the equations that govern the behavior of those objects, $\mathbf{F} = m\mathbf{a}$, for example, are agreed upon by all.

Precisely what constitutes a permissible reference frame depends on the physical law being described, and under what domain it is claimed to hold true. We shall take a break from electromagnetic theory in much of this chapter to discuss frames of reference and the transformations that allow state information (e.g., position, direction, speed) from one frame to be converted to another. The reasons for doing this are both practical and theoretical. Practical applications include the understanding and compensation of real-world phenomenon such as Doppler effects, stellar aberrations, and (in the noninertial case) derived forces such as the centrifugal and Coriolis forces. Theoretically, checking our proposed physical laws against the expectation of reference-frame independence can guide us in the proper format and interpretation of those laws, as this chapter will show.

1 Contrary to colloquial thought, the original concept of relativity was not invented by Einstein [1, 2]. Galileo, for example, discussed it at great length in his *Dialogue Concerning the Two Chief World Systems* in 1632 to combat arguments made against the Copernican model of the solar system, whose opponents claimed that objects could not move the way they seemed to if the Earth were in motion orbiting the Sun.

We will revisit Maxwell's equations only briefly to discuss why they create such a problem for this concept in inertial reference frames, before deriving a new transformation, the Lorentz boost, which preserves the principle of relativity under Maxwell's equations by redefining our basic understanding of space and time.

2.1 GALILEAN TRANSFORMATION

Converting state information from one reference frame to another is very much like changing coordinate systems. Clearly, physical laws should work regardless of whether we use Cartesian coordinates, cylindrical coordinates, or spherical coordinates.[2] Other permissible changes in coordinates that seemed obvious to the likes of Galileo and Newton included translation of the point of origin, rotation about an axis, and relative motion at a constant rate of speed and direction (like our cup of coffee on the train). In this section, we will begin to assemble a mathematical framework with which to describe these transformations.

2.1.1 Translation (Spatial and Temporal)

It goes almost without saying that physical laws should not depend on absolute time, position, or orientation in space; an experiment conducted today facing south in Charlottesville is expected to yield the same result if it is repeated tomorrow facing east in Pasadena. The point of writing down a physical law, to the scientist, is to make predictions about how nature will behave under certain conditions, and, to the engineer, is to enable him or her to recreate those conditions necessary to achieve a desired result. Laws that do not obey at least these simple constraints would not be very useful.

Translation of spatial coordinates amounts to the simple addition of an offset to each Cartesian axis, x, y, and z,

$$x' = x + \Delta x \tag{2.1a}$$

$$y' = y + \Delta y \tag{2.1b}$$

2 However, their forms may change if our notation is not sufficiently general, for example, if we were to explicitly write $\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0$ instead of the more general $\boldsymbol{\nabla} \cdot \mathbf{B} = 0$. The latter is what is called a coordinate-free representation. Our development of relativistic electromagnetic theory will follow a similar progression, starting in Chapter 4 with a form known as tensor analysis that is bound to a selected reference frame and coordinates, only to be supplanted later by a coordinate-free representation in Chapter 5 called spacetime algebra.

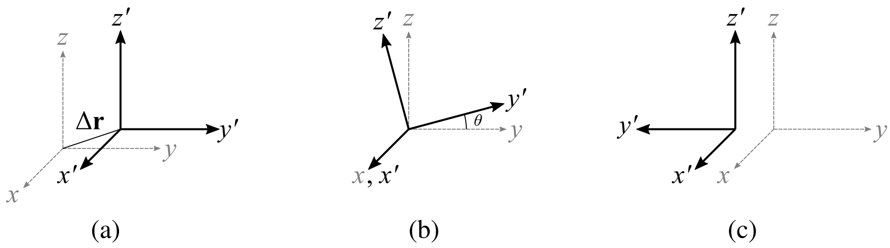


Figure 2.1 Spatial transformations. (a) Translation. (b) Rotation. (c) Reflection.

$$z' = z + \Delta z \tag{2.1c}$$

or, in vector form,

$$\mathbf{r}' = \mathbf{r} + \Delta\mathbf{r} \tag{2.2}$$

as illustrated in Figure 2.1(a). The same general form follows for a shift in time,

$$t' = t + \Delta t \tag{2.3}$$

It will facilitate our discussion later on if we gather our *spatial* and *temporal* coordinates together into a single four-dimensional vector, $\mathbf{X}$, where

$$\mathbf{X} = \begin{pmatrix} ct \\ \mathbf{r} \end{pmatrix} = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \tag{2.4}$$

The vector $\mathbf{X}$ is sometimes called the *four-position*. Note that, to make the units work out, we have scaled the time coordinate by a constant factor, c, having dimensions of speed (say, meters per second). No requirement is made at this point about what c is, only that it is a constant, independent of reference frame. As we shall see eventually in Section 2.3, the speed of light is an especially apropos choice. Vectors such as $\mathbf{X}$ shall be referred to herein as *four-vectors*, to distinguish them from the ordinary three-dimensional spatial vectors with which we are already familiar (the latter will sometimes be called *three-vectors* for clarity).

A general translation of both space and time may therefore be written as

$$\mathbf{X}' = \mathbf{X} + \Delta\mathbf{X} = \begin{pmatrix} ct \\ \mathbf{r} \end{pmatrix} + \begin{pmatrix} c\Delta t \\ \Delta\mathbf{r} \end{pmatrix} = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} + \begin{pmatrix} c\Delta t \\ \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} \tag{2.5}$$

Transformations such as these, involving the addition of four-vectors, are called inhomogeneous.

2.1.2 Rotation

Homogeneous transformations, in contrast, take the form of matrix multiplication. The canonical example is a rotation through a fixed angle about a given axis. Say we wish to rotate our coordinates by an angle θ about the x axis, as shown in Figure 2.1(b). Calculation of the new coordinates is a matter of simple trigonometry,

$$x' = x \tag{2.6a}$$

$$y' = y\cos\theta + z\sin\theta \tag{2.6b}$$

$$z' = z\cos\theta - y\sin\theta \tag{2.6c}$$

and of course $t = t'$. In terms of the four-vector,

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\theta & \sin\theta \\ 0 & 0 & -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \tag{2.7}$$

or $\mathbf{X}' = \mathbf{R}\mathbf{X}$ where

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\theta & \sin\theta \\ 0 & 0 & -\sin\theta & \cos\theta \end{pmatrix} \tag{2.8}$$

Rotations about other Cartesian axes may be derived in a similar manner, or by interchanging the appropriate rows and columns in the matrix, $\mathbf{R}$. Rotation about arbitrary, non-Cartesian axes may be written as the product of two or more Cartesian rotations. In all cases, the resulting transformation matrix is orthogonal [3, 4], meaning that $\mathbf{R}^T = \mathbf{R}^{-1}$.

Neither of the transformations that we have discussed so far require any change to the values of the fields or source quantities in Maxwell's equations, the continuity equation, the Lorentz force law, or any of the constituent definitions. The coordinates x, y, z, and t do not appear explicitly anywhere in those equations, except as variables of differentiation. Derivatives with respect to time are not affected by the addition of an offset, Δt, and the vector spatial derivatives (divergence, gradient, and curl) are unaffected by either offsets or rotation. Take, for example, the del operator in primed coordinates after an x-axis rotation,

$$\nabla' = \mathbf{x}'\frac{\partial}{\partial x'} + \mathbf{y}'\frac{\partial}{\partial y'} + \mathbf{z}'\frac{\partial}{\partial z'} \tag{2.9a}$$

$$= \mathbf{x}\frac{\partial}{\partial x} + (\mathbf{y}\cos\theta + \mathbf{z}\sin\theta)\left(\cos\theta\frac{\partial}{\partial y} + \sin\theta\frac{\partial}{\partial z}\right) + (\mathbf{z}\cos\theta - \mathbf{y}\sin\theta)\left(\cos\theta\frac{\partial}{\partial z} - \sin\theta\frac{\partial}{\partial y}\right) \tag{2.9b}$$

$$= \mathbf{x}\frac{\partial}{\partial x} + \mathbf{y}\frac{\partial}{\partial y} + \mathbf{z}\frac{\partial}{\partial z} = \nabla \tag{2.9c}$$

Thus, for these transformations, when we convert from one reference frame to another, the labels of the coordinates change, but the values of the field parameters do not. If the magnetic field is 7 A/m at some point in one reference frame, then it is 7 A/m at the same point in the other reference frame, although their coordinates will be different. In other words,

$$\mathbf{H}(x', y', z') = \mathbf{H}(x, y, z) \tag{2.10}$$

However, there are some transformations that do cause observers in one reference frame to measure a different value for the same state parameter as observers in another reference frame. The cup of coffee on the train is one example; passengers not only assign it different coordinates than those outside the train, but they perceive it as having a different velocity as well. Something similar will be true of the magnetic field in the next transformation we describe, called a reflection.

2.1.3 Reflection

Note that the determinant of the transformation matrix for the x-axis rotation in (2.8) is unity, or $\det \mathbf{R} = 1$. This is a characteristic of all transformations classified as *proper rotations*.

An *improper rotation* or *improper transformation* is one for which $\det \mathbf{R} = -1$. The inversion of the y-axis coordinate shown in Figure 2.1(c) is one example, and its transformation matrix can be written

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.11}$$

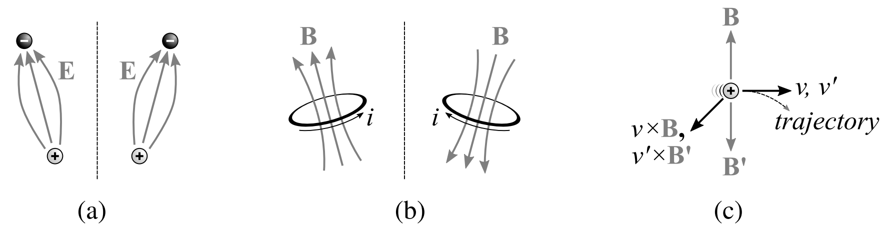


Figure 2.2 Illustration that $\mathbf{E}$ is a polar vector while $\mathbf{B}$ is an axial vector. (a) If the source charge density is mirrored, the electric field is mirrored as well. (b) If the source current density is mirrored, the magnetic flux density is not only mirrored, but also negated. (c) Changing the handedness of the coordinate system (and of all cross products) requires changing the sign of the magnetic flux density so that the Lorentz force law remains true.

By convention, we prefer to set up our coordinate axes so that the z-axis points in the direction of our thumb if we curl the fingers of our right hand from x to y; coordinate systems that obey this convention are said to be right-handed. The transformation in this case reverses that handedness. Clearly, there is nothing special about right-handed coordinates; the laws of physics should apply whether our axes are right-handed or left-handed, but we must be very careful about how we define curls and cross products. Do we consider the right-hand rule an integral part of their definition, or are they defined more formulaically as in (B.4) and (B.17), in which case the results of those operations change handedness to match that of the coordinate system? I am not going to argue ideologically about which of these two is the correct definition; I will only say that there is a subtle difference in the way that $\mathbf{E}$ and $\mathbf{B}$ behave as vectors which is best illustrated if we adopt the latter approach.

Forgetting about coordinate systems for the moment, consider geometrically the two scenarios shown in Figures 2.2(a) and (b). In Figure 2.2(a), we first picture separated positive and negative charges, and the electric field that they create between them, then we draw the mirror image. The electric field is mirrored just as the charges are. On the other hand, if we picture a loop of current creating a magnetic flux, as in Figure 2.2(b), simply mirroring both the current and the magnetic flux does not result in a picture that is accurate with physics. We must additionally reverse the direction of the magnetic flux (from pointing up to pointing down in the figure) in order to be consistent with Maxwell's equations.

Figures 2.2(a) and (b) are examples of active transformation, where the base coordinate axes are not changed, but the source and field distributions are. A passive transformation is illustrated in Figure 2.2(c), wherein the sources are fixed, but the coordinate labels are changed. Here, we have in our initial right-handed coordinate system a positively charged particle moving to the right with velocity v. There is also a vertically oriented magnetic flux density, $\mathbf{B}$. The Lorentz force ($= q\mathbf{v} \times \mathbf{B}$), in accordance with the right-hand rule, is thus directed out of the page, so that the future trajectory of the particle bends toward us. If we convert our coordinate system to a left-handed one, the observable velocity ($\mathbf{v}'$) is still rightward, but the cross product ($\mathbf{v}' \times \mathbf{B}'$) is mathematically left-handed, so consistency with the Lorentz force requires that the magnetic flux density have the opposite orientation. Thus, observers using a left-handed coordinate system will see a magnetic field having the opposite sign of those using a right-handed coordinate system.

$\mathbf{E}$ is what is known as a polar vector [5]. Polar vectors match their own mirror images upon active reflection. $\mathbf{B}$ is different; $\mathbf{B}$ is defined only in terms of its interaction with moving particles through the cross product, and with the electric field through its curl. $\mathbf{B}$ is known as an axial vector or pseudovector, one that must be flipped upon active reflection. Another example of an axial vector from basic mechanics is angular momentum.

Much will be made of the distinction between polar vectors and axial vectors in the mathematics of special relativity, so it is important to have an intuitive understanding of the difference. Polar vectors represent a direction along which some action occurs — for example, the electric field, as a polar vector, indicates the direction of acceleration of a charged particle in the presence of that field. In contrast, axial vectors define an axis around which the action occurs; the magnetic field, an axial vector, does not directly indicate the direction of acceleration, but rather the axis around which that acceleration occurs. It is no accident that axial vectors are virtually always defined via cross products (or curls, when differentiation is involved), thus necessarily invoking the right-hand rule to determine the actual direction of motion.

Why does any of that matter? For one, it proves that $\mathbf{E}$ and $\mathbf{B}$ cannot both transform the way other four-vectors do — that is, we cannot write

$$\begin{pmatrix} ? \\ E'_x \\ E'_y \\ E'_z \end{pmatrix} = \mathbf{R} \begin{pmatrix} ? \\ E_x \\ E_y \\ E_z \end{pmatrix} \tag{2.12}$$

and

$$\begin{pmatrix} ? \\ B'_x \\ B'_y \\ B'_z \end{pmatrix} = \mathbf{R} \begin{pmatrix} ? \\ B_x \\ B_y \\ B_z \end{pmatrix} \tag{2.13}$$

Setting aside what the first, timelike component in each four-vector should actually be in this hypothetical scenario, the equations above would either imply that both $\mathbf{E}$ and $\mathbf{B}$ will flip direction, or that neither will. Running both field vectors through this operation results in a picture that, while mathematically acceptable, is not consistent with Maxwell's equations.

We will have more to say in later chapters about how to correctly transform $\mathbf{E}$ and $\mathbf{B}$ into different reference frames. Before we leave this section, however, there is one more thing we can say about other quantities we might wish to transform. Ultimately, this point stems from the fact that we require our physical laws to work for all valid coordinate systems, whether left-handed or right-handed; thus, no physical law can equate a polar vector to an axial vector.

First, we note that simple scalars, like mass and electric charge, do not spontaneously change sign upon reflection.[3] This means that since ε_0 and μ_0 are scalars, $\mathbf{D} = \varepsilon_0\mathbf{E}$ must be a polar vector like $\mathbf{E}$, while $\mathbf{H} = \mu_0^{-1}\mathbf{B}$ is an axial vector. Further, the Lorentz force law in the electrostatic case, $\mathbf{F} = q\mathbf{E}$, implies that since $\mathbf{E}$ is a polar vector, force is also a polar vector. Newton's law of motion, $\mathbf{F} = m\mathbf{a}$, then shows that acceleration, $\mathbf{a}$, is a polar vector, as is velocity, $\mathbf{v}$, its integral.

The chain of reasoning continues in this manner, until we have classified all vectors of interest as polar or axial. An interesting special case is that of a hypothetical magnetic charge. Although none has ever been observed in nature, there is nothing about Maxwell's equations or other physical theories that forbid it.[4] It would appear as a source term in the divergence of magnetic flux, $\nabla \cdot \mathbf{B} = \rho_m$. Since $\mathbf{B}$ is an axial vector, the magnetic charge, ρ_m, would have to change sign upon reflection. It would be what we call a *pseudoscalar*. By similar reasoning, the magnetic current, which would have to be added to Faraday's law, would be classified as an axial vector. A summary of these classifications is given in Table 2.1.

3 In the case of electric charge, this is actually more of a convention than an absolute truth. Since electric forces depend only on the product of two charges, if both changed sign under reflective coordinate transformations (however strange that may seem), the end result would be the same. Much of the logical deductions which follow would be reversed as a consequence, but the resulting physics would still be internally consistent.

4 On the contrary, Paul Dirac famously showed that if even a single particle of magnetic charge existed somewhere in the universe, it would provide a satisfying explanation for the observed quantization of electric charge [6].

Table 2.1
Classification of Vectors and Scalars

Name	Symbol	Classification
Electric field and displacement	$\mathbf{E}$, $\mathbf{D}$	polar vectors
Magnetic field and flux density	$\mathbf{H}$, $\mathbf{B}$	axial vectors
Permittivity and permeability	ε_0, μ_0	scalars
Mass, charge, and charge density	m, q, ρ	scalars
Magnetic charge and charge density	q_m, ρ_m	pseudoscalars
Current density	$\mathbf{J}$	polar vector
Force, velocity, and acceleration	$\mathbf{F}$, $\mathbf{v}$, $\mathbf{a}$	polar vectors
Angular momentum	$\mathbf{L}$	axial vector

2.1.4 Boosts

The last transformation we will talk about is the one we actually started the chapter with, in the example of the coffee cup on the train. This type of transformation is called a boost, wherein one reference frame is in motion at a constant velocity relative to another reference frame, which for convenience we shall call the *rest frame*.

The stipulation that the velocity of relative motion must be constant means that both will be *inertial reference frames*. The concept of an inertial reference frame was well known to both Newton and Galileo. Newton's laws (at least at velocities much less than the speed of light) are the same in all inertial reference frames. As anyone who has ridden a train or a plane knows, the vehicles in which they are traveling may jostle them about when the track curves or the air is turbulent, but at least when the ride is smooth and straight one can stand, walk, perform cartwheels, or juggle balls (at least as well as they could before) without having to relearn the physics.

However, Newton's laws do change in noninertial reference frames. A child who rolls a ball on a merry-go-around will see it move in strange and unfamiliar ways (from his perspective) under the apparent influence of the "fictitious" centrifugal and Coriolis forces. In relativistic terms, the difference between inertial reference frames and noninertial reference frames is the difference between special relativity and general relativity. For the purposes of this book, we shall always restrict ourselves to inertial reference frames, or to special relativity.

Example: In Galileo's view, the equations relating the unprimed coordinates of the rest frame to the primed coordinates of the moving frame, for motion in the x direction, are

$$t' = t \tag{2.14a}$$

$$x' = x - vt \tag{2.14b}$$

In terms of four-vectors, a boost of velocity v in the x direction is $\mathbf{X}' = \mathbf{G}\mathbf{X}$, where

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\beta & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.15}$$

and where $\beta = v/c$ is the normalized velocity of motion. Be reminded that the constant c appears here only because of the way we chose to scale the time component of our four-vectors; there is nothing special about its value, and no requirement that it be the speed of light (when it is normalized to the speed of light, β is sometimes called the *Jackson number*). Coordinates in the moving reference frame are therefore calculated as follows

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\beta & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} ct \\ x - vt \\ y \\ z \end{pmatrix} \tag{2.16}$$

Thus, if persons on the ground perceive the coffee cup as moving to the right at the speed of the train, $x = vt$, then passengers on the train will perceive its position as

$$x' = x - vt = vt - vt = 0 \tag{2.17}$$

In other words, to passengers on the train, the coffee cup is at rest. More generally, the velocity of any moving object observed on and off the train are related by

$$\frac{dx'}{dt} = \frac{dx}{dt} - v \tag{2.18a}$$

$$\therefore \frac{dx}{dt} = \frac{dx'}{dt} + v \tag{2.18b}$$

Thus, if a passenger on the train throws an object forward (toward the front of the train), observers on the ground see its velocity added to that of the train. Conversely, if the object is thrown backward, they see its velocity subtracted from that of the train.

2.2 SPACETIME

We now come to the great contradiction which so puzzled scientists in the latter part of the nineteenth century: the independence of the speed of light from the reference frame of a moving observer. Resolution of this apparent conflict required changing the way we transform between inertial reference frames, and the interpretation of that formula is what we now call special relativity.

2.2.1 Invariability of the Speed of Light

Recall the manner in which the propagation speed of electromagnetic waves was calculated (not measured) by Maxwell. Purely as a consequence of its mathematical structure, the speed of light in a vacuum had to be

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} \tag{2.19}$$

The values of the vacuum permittivity and permeability (ε_0 and μ_0) that fed into this equation were, in turn, measured by experiments like Coulomb's and Ampère's, experiments that focused on quantifying the intensity of fundamental forces, not measuring actual speeds.

It is important to remember that the permeability and permittivity of physical materials, like glass or other insulators, through which light and other electromagnetic waves may pass, are phenomenological factors that average the effects of large numbers of molecules in order to present a simpler picture for the macroscopic behavior of the $\mathbf{E}$ and $\mathbf{H}$ fields, as described in Section 1.4.1. They are created by the constrained movements of countless individual charges (electrons and atomic nuclei) in response to external stimuli — fields that in reality exist in the vacuum spanning the tiny spaces between those charges. By subsuming the combined influence of these discrete charges into proportionality constants, ε_r and μ_r, we avoid the intractable problem of tracking the movements of all such bound charges as individual terms of ρ and $\mathbf{J}$. Only those sources that are not bound to the material structure, known as free charges and currents, need to be included in specific solutions of Maxwell's equations.

Presumably, if the medium were in motion, then a Galilean boost transformation could be applied to each and every molecule and charge individually, subject to the constraining forces binding them to the material structure. Their net effect could then be averaged over once again to create new effective constituent parameters — let's call them ε' and μ' — which in turn might show a different speed of wave propagation. Indeed, experiments have confirmed that light passing through a material in motion is dragged to some extent by the medium, but not to the degree that nonrelativistic analysis would predict [7].

In a vacuum, however, there are no electrons or atomic nuclei to contribute to the constituent parameters, ε_0 and μ_0. There is no medium to which we can assign a velocity relative to the observer. In a vacuum, these parameters have more the character of unit-conversion factors, like millimeters per inch or grams per pound. Put simply, they are fundamental constants underpinning the SI unit system,[5] not phenomenological manifestations of countless molecular charges working in concert.

Nevertheless, to most scientists of the nineteenth century (and many still into the twentieth century), the basic truth of (2.14), the Galilean transformation of motion, seemed just too obvious to be wrong. Maxwell's equations felt abstract in comparison, and it was thus assumed that Maxwell had missed something (after all, the initial experiments that led to Maxwell's equations did miss the displacement current). Perhaps there was an underlying medium which led to the emergence of ε_0 and μ_0, whose velocity we could measure and account for in our equations. That medium was to be known as the *luminiferous aether* [1].

Conceptually, at least, we could imagine resurrecting Coulomb, Ampère, Faraday, and the rest, and put them all onto a fast-moving train to repeat their historic experiments. If conducted with sufficient precision (far greater than those scientists could actually achieve given the instruments of their day), we might expect to measure slightly different values for μ_0 and ε_0, such that the speed of light we then calculated would differ from the original result by an amount commensurate with the speed of their moving platform.

Should the reader have any doubt as to the outcomes of those hypothetical experiments, it may help to keep in mind that the exercise has, in a way, already been done! All of these scientists, and those that followed them who confirmed

---

5 See Appendix A for a unit system in which these parameters are not needed or, equivalently, have a value of 1.

their results again and again with ever-increasing precision, carried out their measurements on a planet whose surface is rotating many hundreds to a thousand miles per hour (depending on the latitude), and orbiting at 67,000 miles per hour around a star that is in motion around the galactic center at a speed of roughly half a million miles per hour [8].[6] The very fact that we cannot tell that we are living out our lives on a fast-moving platform is exactly the reason Galileo postulated his principle of relativity in the first place.

In reality, these historical experiments were hindered by a lack of precision. The speed of light is incredibly fast, so the impact of our comparatively sluggish motion is expected to be quite small. New experiments were devised, most famously the Michelson-Morley experiment [9], which attempted to measure the motion of the Earth through the aether by directly observing changes in the speed of light and other indirect effects. I will not go into these experiments in detail, other than to say that they were designed to have precision more than sufficient to detect even the slower motions of the Earth described above. Had there been any directional bias or seasonal variation of the speed of light associated with the Earth's motions, we would surely have detected it.

The continuing failure of these experiments to detect the reference frame of the aether in any form forced the scientific community to assign it ever more exotic properties — such as the ability to flow smoothly through solid materials like glass while being dragged along with the Earth in its orbit around the Sun. Even worse, measurements on the drag coefficient of light passing through transparent materials (an experiment that will be described in more detail in Section 6.2.4) seemed to indicate that its drag coefficient had a dependence on refractive index, which generally varies according to wavelength and sometimes polarization — suggesting that the aether could move at two different velocities at the same time [7]! God, it seemed, had perversely designed the aether specifically to foil our attempts at detecting it.

2.2.2 The Loss of Simultaneity

It was Albert Einstein who ultimately abandoned this approach. Proceeding logically from the assumption that the Galilean principle of relativity (the idea that physics should be no different in any inertial reference frame) as well as Maxwell's equations were both correct, he conducted a thought experiment that effectively

6 I once heard someone jokingly ask: If he were indeed moving through space at such a dizzying rate of speed, then why didn't his hair get messed up?

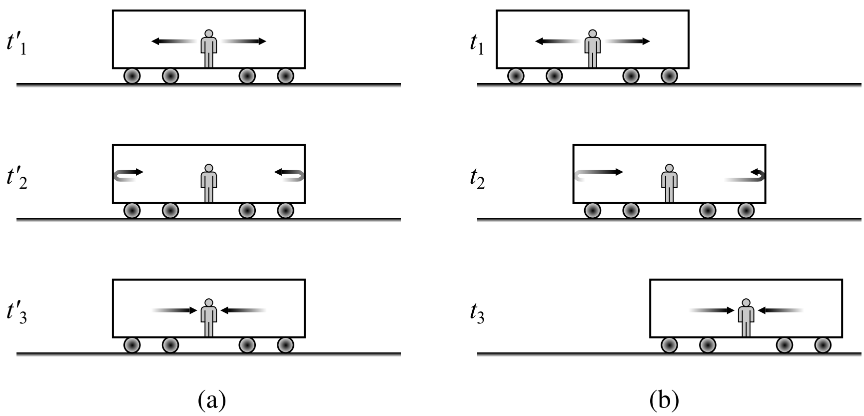


Figure 2.3 Illustration of the loss of simultaneity, using pulses of light inside a rail car with reflective mirrors at both ends. (a) Sequence of events as observed by the passenger, in which the two pulses of light reflect off the mirrors at the same instant. (b) Sequence of events as seen by an external bystander, in which the backward-traveling pulse reflects first, and the forward-traveling pulse reflects second.

redefined what it meant to measure distance and time. I will adapt his scenario slightly in what follows to suit my own narrative style.

Imagine you are standing in the center of a rail car. (I will not tell you at this point whether the car is moving; remember that, from your viewpoint inside the car, you cannot tell, nor does it matter.) You have carefully measured the distance from your position in the center of the car to both ends, where you have placed two mirrors. You have ensured that your distance to either of them is the same.

From that central position, you set off a flash bulb, sending pulses of light in either direction at speed c, as shown in Figure 2.3(a). Let us call this moment t'_1. At time t'_2, both pulses reflect off of the mirrors at the ends of the car simultaneously. Finally, at time t'_3, the pulses return to the center, arriving back at the exact same instant. From your viewpoint, the reflection of both pulses of light off the mirrors are simultaneous events.

Now consider how the situation would appear to a bystander outside the rail car, with the added assumption that the car is moving to the right (from the bystander's perspective) at some velocity, v. As this is what we would normally call the rest frame, we will use unprimed time coordinates. At time t_1, the flash bulb goes off, sending the beams of light once again to the right and left at speed c, dictated by Maxwell's equations. In this case, however, the back end of the car is

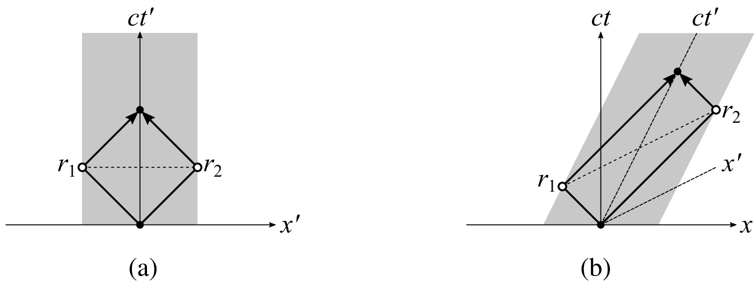


Figure 2.4 Minkowski diagrams for the rail car thought experiment in Figure 2.3. (a) The passenger's reference frame. The extent of the rail car is shown by the shaded region. Reflection events r_1 and r_2 occur simultaneously. (b) The bystander's reference frame. The dashed line labeled ct' is the passenger's worldline in this reference frame. The dashed line labeled x' is a contour of simultaneity for the passenger. Reflection event r_1 appears to precede r_2 from the bystander's perspective, while they are simultaneous (parallel to the x' axis) in the passenger's reference frame. We make no claim as to the relative scales of the primed and unprimed axes at this point.

moving toward the backward-traveling light pulse, causing it to reflect first, before the forward-traveling pulse connects with its mirror, which is moving away from it. Finally, at time t_3, both pulses return to you, the passenger at the center of the car, at the same instant, but at a position some distance to the right from where you started.

2.2.3 Minkowski Space

It is useful to illustrate scenarios like this with a plot known as a *Minkowski diagram*, shown in Figure 2.4. We plot one spatial coordinate on the horizontal axis, and time on the vertical axis. For dimensional consistency, let us scale the time coordinate by c, as we did with the four-vectors (only now we will stipulate that c be the speed of light, for convenience, given the arguments about its special nature in the preceding section). Note that this means a beam of light will always follow a 45° angle in the diagram, in any reference frame. We denote the extent of the rail car by a shaded box. In your reference frame as the passenger, Figure 2.4(a), that shaded box is upright, having vertical sides, since the rail car is unmoving in this reference frame. The beams of light, drawn with black arrows, reflect off the ends at event points labeled r_1 and r_2. In the bystander's reference frame, Figure 2.4(b), the shaded box leans to the right, since the ends of the rail car are moving. The rays of light, still drawn at 45° angles, reflect again off the ends of the car, only this time the event

r_1 precedes r_2 as it happens at an earlier time coordinate, ct. Your location as the passenger may be drawn in this reference frame through the center of the rail car as a dashed line labeled ct'. This is called your *worldline*, and has the following form in the bystander's coordinates,

$$x = vt = \beta ct \tag{2.20}$$

where, again, $\beta = v/c$ is the normalized velocity. Also drawn with a dashed line is x', the contour of simultaneity from your point of view at time $t' = 0$. Since the reflection events occur simultaneously to you, this line must be parallel to the line passing through r_1 and r_2. It may be found geometrically as

$$ct = \beta x \tag{2.21}$$

We may conclude from this exercise that events occurring simultaneously at two different locations in one reference frame (such as the reflections r_1 and r_2) may occur at different times in another reference frame. The sequence depends on the reference frame: from the bystander's perspective, r_1 precedes r_2, but we can imagine another reference frame, say that of a passenger on another train moving at $2v$ to the right. To a passenger on that second train, the first rail car is trailing leftward at speed v, and from her perspective, the reflection r_2 precedes r_1. In contrast, it is also apparent that events that occur simultaneously and at the same location in one reference frame (such as the initial emission of the light pulses and their eventual convergence back at the source) are simultaneous in all reference frames.

The reader may be feeling some unease at this point, for it may seem like the reversal of a sequence of events would wreak havoc with our concept of cause-and-effect, or even, for that matter, of free will. However, we will eventually see that the degree to which spatially separated events may be reordered is limited so that causality is always preserved.

It bears remembering why the conclusions we have drawn from this scenario depend on the use of light as a tracer. Had you launched something else from the center of the car, say a pair of tennis balls, not only would things unfold at a much slower pace, but the overall picture in Figure 2.3(b) would be quite different. An external observer would see the tennis ball thrown toward the front of the train departing your hand at a much higher velocity than the one thrown to the back of the train. You would throw them with equal force and give them equal velocities from your own point of view, but the external observer would see that velocity added to and subtracted from the velocity of the train. Had sound waves been used, then the air within the train car, as the supporting structure of the sound waves, would likewise carry them forward at a faster rate from the external viewer's perspective, while retarding those waves that propagate backward. Maxwell's equations tell us that light, uniquely, does not act that way. There is no way to throw light any faster (or slower) than c, nor is there a supporting medium to give it a boost forward or a drag backward. Light travels at the velocity dictated by its nature, regardless of the motion of the platform that launched it or the motion of the one observing it.

2.2.4 Seeing Is Not Believing

I want to take a moment to emphasize that the phenomena we have described are not merely optical illusions brought about by the time-of-flight required for the light-image of events to reach an observer's eyes. I will somewhat loosely refer to what an observer sees to mean what a careful and well-funded experimenter could determine, by marking distances, recording times, and measuring the speeds of various rays of light (assuming he didn't simply trust Maxwell's equations from the outset) to calculate exactly when events must have occurred at specific locations, apart from the moments he actually witnessed them. How a particular experiment appears, visually, to the more casual bystander is a different question, one that would require much more complicated diagrams than those in Figure 2.4 to account for the viewing angle of the bystander and the travel time needed for the light-evidence of various events to reach him.

2.3 LORENTZ TRANSFORMATION

Let us now apply some mathematical rigor to the conceptual picture we have just created [10]. Specifically, we wish to rewrite (2.14), the Galilean boost transformation equations, in a way that is consistent with what we have learned about light.

2.3.1 Derivation

We may constrain their form as follows. First, (2.20) tells us that the worldline of an observer in motion, characterized by $x' = 0$, is given by $x = \beta ct$. To ensure that this is the case, we may pull out a factor of $x - \beta ct$ from the equation for x',

$$x' = (x - \beta ct)\gamma(x, t, \beta) \tag{2.22}$$

where $\gamma(x,t,\beta)$ is some yet-to-be-determined function of space, time, and inertial velocity. The expected independence of things like relative speed from the coordinate systems' points of origin and reference time further shows that the term γ can be neither a function of x nor of t,

$$x' = (x - \beta ct)\gamma(\beta) \tag{2.23}$$

Finally, the expected symmetry with respect to motion in the left and right directions further implies that γ is an even function of β, or $\gamma(\beta) = \gamma(-\beta)$, and we may write the inverse transformation as

$$x = (x' + \beta ct')\gamma(\beta) \tag{2.24}$$

In a similar manner, we may conclude from (2.21) that $ct = \beta x$ whenever $t' = 0$, or, in other words, that the equation for t' has a factor of $ct - \beta x$ in it,

$$ct' = (ct - \beta x)\kappa(\beta) \tag{2.25}$$

where, again from symmetry, $\kappa(\beta) = \kappa(-\beta)$, and

$$ct = (ct' + \beta x')\kappa(\beta) \tag{2.26}$$

Collecting these results into matrix form, we have

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \kappa & -\kappa\beta \\ -\gamma\beta & \gamma \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix} \tag{2.27a}$$

$$\begin{pmatrix} ct \\ x \end{pmatrix} = \begin{pmatrix} \kappa & \kappa\beta \\ \gamma\beta & \gamma \end{pmatrix} \begin{pmatrix} ct' \\ x' \end{pmatrix} \tag{2.27b}$$

Therefore,

$$\begin{pmatrix} \kappa & \kappa\beta \\ \gamma\beta & \gamma \end{pmatrix} = \begin{pmatrix} \kappa & -\kappa\beta \\ -\gamma\beta & \gamma \end{pmatrix}^{-1} = \frac{1}{\gamma\kappa\left(1-\beta^2\right)} \begin{pmatrix} \gamma & \kappa\beta \\ \gamma\beta & \kappa \end{pmatrix} \tag{2.28a}$$

$$= \frac{1}{1-\beta^2} \begin{pmatrix} \kappa^{-1} & \gamma^{-1}\beta \\ \kappa^{-1}\beta & \gamma^{-1} \end{pmatrix} \tag{2.28b}$$

or

$$\gamma(\beta) = \kappa(\beta) = \frac{1}{\sqrt{1-\beta^2}} \tag{2.29}$$

The term γ is known as the *Lorentz factor*. We thus have for the Lorentz transformation,

$$ct' = (ct - \beta x)\gamma \tag{2.30a}$$

$$x' = (x - \beta ct)\gamma \tag{2.30b}$$

$$y' = y \tag{2.30c}$$

$$z' = z \tag{2.30d}$$

or, in terms of the complete four-vector,

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \tag{2.31}$$

More simply, $\mathbf{X}' = \mathbf{L}\mathbf{X}$, where

$$\mathbf{L} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.32}$$

A plot of the key factors in the Lorentz boost transformation as a function of relative velocity is in Figure 2.5(a).

Like proper rotations, the Lorentz boost operator has a determinant of unity,

$$\det \mathbf{L} = \gamma^2 - \beta^2\gamma^2 = \frac{1 - \beta^2}{1 - \beta^2} = 1 \tag{2.33}$$

Given the reciprocal nature of the principle of relativity, we should expect the inverse of $\mathbf{L}$ to look very similar, except for a sign change of β,

$$\mathbf{L}^{-1}(\beta) = \frac{\text{adj}(\mathbf{L})}{\det \mathbf{L}} = \begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & \gamma^2 - \beta^2\gamma^2 & 0 \\ 0 & 0 & 0 & \gamma^2 - \beta^2\gamma^2 \end{pmatrix} \tag{2.34a}$$

$$= \begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \mathbf{L}(-\beta) \tag{2.34b}$$

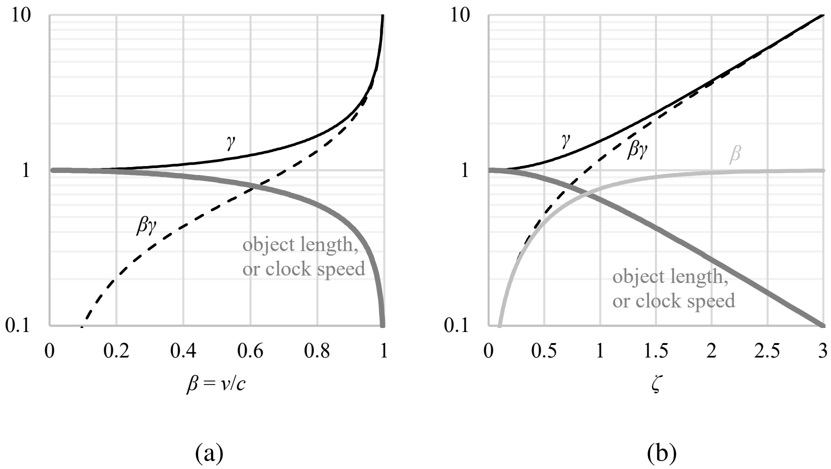


Figure 2.5 Lorentz transformation factors as a function of (a) relative velocity, β, and (b) rapidity, ζ.

Lorentzian boosts along other Cartesian axes may be found by exchanging the appropriate rows and columns, or, for directions not aligned to a particular axis, by the combined transformations (that is, matrix products) corresponding to a rotation, a boost, and a (de)rotation, in that order.

Alternatively, we can associate the x coordinate in the above derivation with a position in arbitrary coordinates, $\mathbf{r}$, aligned with the direction of motion, $\mathbf{n}$,

$$x \rightarrow \mathbf{r} \cdot \mathbf{n} \tag{2.35}$$

The remaining spatial coordinates may then be written as the residual,

$$y\mathbf{y} + z\mathbf{z} \rightarrow \mathbf{r} - (\mathbf{r} \cdot \mathbf{n})\mathbf{n} \tag{2.36}$$

Applying this to (2.30), we have

$$ct' = (ct - \beta\mathbf{r} \cdot \mathbf{n})\gamma \tag{2.37a}$$

$$\mathbf{r}' \cdot \mathbf{n} = (\mathbf{r} \cdot \mathbf{n} - \beta ct)\gamma \tag{2.37b}$$

$$\mathbf{r}' - (\mathbf{r}' \cdot \mathbf{n})\mathbf{n} = \mathbf{r} - (\mathbf{r} \cdot \mathbf{n})\mathbf{n} \tag{2.37c}$$

or

$$ct' = (ct - \mathbf{r} \cdot \boldsymbol{\beta})\gamma \tag{2.38a}$$

$$\mathbf{r}' = \mathbf{r} + (\gamma - 1)(\mathbf{r} \cdot \mathbf{n})\mathbf{n} - \gamma c t \boldsymbol{\beta} \tag{2.38b}$$

where

$$\boldsymbol{\beta} = \beta \mathbf{n} \tag{2.39a}$$

$$= \beta \left(n_x \mathbf{x} + n_y \mathbf{y} + n_z \mathbf{z} \right) \tag{2.39b}$$

$$= \beta_x \mathbf{x} + \beta_y \mathbf{y} + \beta_z \mathbf{z} \tag{2.39c}$$

Equation (2.38) represents a Lorentz boost transformation in arbitrary vector notation. We may expand these expressions back into matrix form for a Lorentz boost in any direction [11],

$$\mathbf{L} = \begin{pmatrix} \gamma & -\beta_x \gamma & -\beta_y \gamma & -\beta_z \gamma \\ -\beta_x \gamma & 1 + (\gamma - 1) n_x^2 & (\gamma - 1) n_x n_y & (\gamma - 1) n_x n_z \\ -\beta_y \gamma & (\gamma - 1) n_y n_x & 1 + (\gamma - 1) n_y^2 & (\gamma - 1) n_y n_z \\ -\beta_z \gamma & (\gamma - 1) n_z n_x & (\gamma - 1) n_z n_y & 1 + (\gamma - 1) n_z^2 \end{pmatrix} \tag{2.40}$$

2.3.2 Rapidity

Note that the Lorentz factor, γ, becomes infinite when $\beta = \pm 1$, corresponding to an inertial reference frame moving at the speed of light with respect to the rest frame. In fact, mechanical analysis would show that it would take an infinite amount of energy to accelerate any mass-bearing object or observer up to that speed. We will thus take it as a constraint that $|\beta| < 1$. In that regime, we must always have $\gamma > 1$.

It can be useful to capture this limiting behavior by defining β in terms of a new parameter, ζ, known as the rapidity, where

$$\beta = \tanh \zeta \tag{2.41a}$$

$$\therefore \gamma = \frac{1}{\sqrt{1 - \tanh^2 \zeta}} = \cosh \zeta \tag{2.41b}$$

$$\therefore \beta \gamma = \tanh \zeta \cosh \zeta = \sinh \zeta \tag{2.41c}$$

The nature of the hyperbolic tangent function ensures that $-1 < \beta < 1$ for all $-\infty < \zeta < \infty$. It also puts the Lorentz boost operator in a form that closely resembles proper rotations in normal space, but instead using the hyperbolic functions,

$$\mathbf{L}(\zeta) = \begin{pmatrix} \cosh \zeta & -\sinh \zeta & 0 & 0 \\ -\sinh \zeta & \cosh \zeta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.42}$$

Table 2.2
Lorentz Boost Formulas

Description	Formula
Normalized, x-directed	$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}$
Rapidity, x-directed	$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cosh\zeta & -\sinh\zeta & 0 & 0 \\ -\sinh\zeta & \cosh\zeta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}$
General	$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\beta_x\gamma & -\beta_y\gamma & -\beta_z\gamma \\ -\beta_x\gamma & 1+(\gamma-1)n_x^2 & (\gamma-1)n_xn_y & (\gamma-1)n_xn_z \\ -\beta_y\gamma & (\gamma-1)n_yn_x & 1+(\gamma-1)n_y^2 & (\gamma-1)n_yn_z \\ -\beta_z\gamma & (\gamma-1)n_zn_x & (\gamma-1)n_zn_y & 1+(\gamma-1)n_z^2 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}$
Block	$\begin{pmatrix} ct' \\ \mathbf{r}' \end{pmatrix} = \begin{pmatrix} \gamma & -\boldsymbol{\beta}^T\gamma \\ -\boldsymbol{\beta}\gamma & \mathbf{L}_s \end{pmatrix} \begin{pmatrix} ct \\ \mathbf{r} \end{pmatrix}$
Vectorized	$ct' = (ct - \mathbf{r}\cdot\boldsymbol{\beta})\gamma$ $\mathbf{r}' = \mathbf{r} + (\gamma-1)(\mathbf{r}\cdot\mathbf{n})\mathbf{n} - \gamma ct\boldsymbol{\beta}$

where, for illustrative purposes, a boost in the x direction has been described. (This new parameterization will also be convenient notation to use in electromagnetic work where γ and β are already well-established symbols for the propagation and phase constants of guided waves, respectively.) The Lorentz transformation factors are thus shown in terms of the rapidity parameter in Figure 2.5(b).

A summary of Lorentz boost expressions in various forms is given in Table 2.2.

2.3.3 Length Contraction and Time Dilation

The Lorentz transformation is a linear function of spacetime coordinates. Among other things, this means that differences in four-positions transform in the same way as absolute four-positions,

$$\Delta \mathbf{X}' = \mathbf{\Lambda} \Delta \mathbf{X} \tag{2.43}$$

where $\mathbf{\Lambda}$ is any rotation or Lorentzian boost.

Suppose we are in the rest frame observing an object moving in the x direction. We measure its length, which is given by the difference in the x coordinates of its endpoints, Δx. Since the endpoints are in motion, we are careful to measure their coordinates at the same instant in time, $\Delta t = 0$. We then apply a Lorentz boost to catch up to the object, so that we may compare the original length, Δx, to its length in a comoving reference frame, $\Delta x'$,

$$\Delta x' = \Delta x \cosh \zeta - \Delta t \sinh \zeta = \Delta x \cosh \zeta \tag{2.44}$$

or

$$\Delta x = \frac{\Delta x'}{\cosh \zeta} \tag{2.45}$$

Evidently, the object was shortened by a factor of $\cosh \zeta$, the Lorentz factor, due to its motion. This effect is known as *Lorentz contraction*. Note that the object's extents in the y and z directions are unchanged.

Alternatively, we may observe the ticking of a clock in our own rest frame at regular intervals, Δt. Since the clock is unmoving, each tick is recorded at the same location, $\Delta x = 0$. We then apply a Lorentz boost to set the clock in motion, and observe the time intervals again,

$$c\Delta t' = c\Delta t \cosh \zeta - \Delta x \sinh \zeta = c\Delta t \cosh \zeta \tag{2.46a}$$

$$\therefore \Delta t' = \Delta t \cosh \zeta \tag{2.46b}$$

The time interval between ticks has apparently increased, again by a factor of $\cosh \zeta$. This is known as *time dilation*.

To summarize, from the viewpoint of a stationary observer, objects in motion are shortened, and clocks in motion run slower. Since neither frame has any preferential status, this effect works in both directions: A moving meterstick is shortened in the stationary reference frame, while a stationary meterstick is shortened from the viewpoint of the moving reference frame. The length contraction and time dilation factor is shown in Figure 2.5 as the curve labeled "object length, or clock speed."

Figure 2.6 Minkowski diagrams with gridlines to scale (e.g., x, ct, x', and ct' are all in meters) for $\beta = \tanh\zeta = 0.6$. (a) In the plane of the rest frame. (b) In the plane of the moving frame. The path taken by a particular event through various Lorentz boosts is hyperbolic, with asymptotes at $45°$, as shown with the dashed lines.

With these equations in hand, we are now able to draw Minkowski diagrams (though perhaps not easily) with grids in both reference frames at the same scale, as shown in Figure 2.6.

As a corollary, we may also say that if an object is moving at velocity $v = c\tanh\zeta$ in the rest frame, then its surroundings are moving past it at the same velocity in its own reference frame, but in the opposite direction, for the perceived distance to a destination point is shortened from the object's point of view, but the time it experiences in getting there is shortened by the same factor. (However, the apparent relative velocity between two objects is not the same in all other reference frames, as will be discussed in Section 3.1.1.)

2.3.4 A Question of Precedence

The Lorentz transformation equations are so named because they were first articulated by Hendrik Antoon Lorentz [12], a contemporary of Einstein's, notably before Einstein formulated the theory of special relativity. However, Lorentz's interpretation of these equations was wholly different from Einstein's. Lorentz thought of length contraction as the physical compression of material objects in response to being pushed at high speed through a viscous fluid, the aether, and moreover that the formula he derived was but the first, largest term in a kind of infinite series describing this nonlinear effect. This first term would have to have had just such a form to fool the measurements of those like Michelson and Morley, but ongoing, more precise measurements, he believed, would eventually determine additional terms and thus settle the absolute reference frame of the aether.[7] Einstein abandoned that hope, and in doing so dismantled our core assumptions about the very natures of space and time. In Lorentz's words, Einstein presented these principles "not [as] a fortuitous compensation of opposing effects, but the manifestation of a general and fundamental principle." While everyone else was asking, "Where did Maxwell's equations go wrong?" Einstein asked, "What if Maxwell's equations are right?"

2.3.5 Spacetime Intervals and Proper Time

It is reasonable to ask at this point if there is any measurable quantity that has the same value in all reference frames. In fact, there are many quantities that are invariant under all Lorentz transformations (known as the *restricted Lorentz group*, or simply the Lorentz group, which for the rest of this book will be meant to include Lorentz boosts, as well as proper rotations, but not reflections). One of the simplest is known as the *proper time* interval, $\Delta\tau$, where

$$(c\Delta\tau)^2 = c^2\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2 \tag{2.47}$$

or, more simply, the proper time in absolute coordinates,

$$(c\tau)^2 = (ct)^2 - x^2 - y^2 - z^2 \tag{2.48}$$

The proper time may be thought of as a kind of *inner product*[8] of four-positions, given by the following formula,

$$(c\tau)^2 = \mathbf{X} \cdot \mathbf{X} = \mathbf{X}^T \boldsymbol{\eta} \mathbf{X} \tag{2.49}$$

where the superscript T indicates transposition, and $\boldsymbol{\eta}$ is called the metric of the Minkowski space, given by

$$\boldsymbol{\eta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -\mathbf{I} \end{pmatrix} \tag{2.50}$$

7 Lorentz himself conceded that this explanation felt a little contrived, but it was the only explanation he could come up with that seemed to fit the facts [13].

8 For convenience, the dot notation, when used with ordinary three-vectors, will represent the standard inner product, while its use with four-vectors shall imply the matrix form shown here, incorporating the metric, $\boldsymbol{\eta}$.

The metric is said to have a signature of $(+---)$, highlighting the signs of the diagonal elements. It is both symmetric and orthogonal, such that $\boldsymbol{\eta} = \boldsymbol{\eta}^T = \boldsymbol{\eta}^{-1}$.

One may show that the proper time is invariant under Lorentz transformations as follows. Consider the proper time in a transformed reference frame,

$$(c\tau')^2 = (\mathbf{X}')^T \boldsymbol{\eta}\mathbf{X}' = (\boldsymbol{\Lambda}\mathbf{X})^T \boldsymbol{\eta}\boldsymbol{\Lambda}\mathbf{X} = \mathbf{X}^T \left(\boldsymbol{\Lambda}^T\boldsymbol{\eta}\boldsymbol{\Lambda}\right)\mathbf{X} \tag{2.51}$$

The matrix $\boldsymbol{\Lambda}$ is meant to represent any possible combination of Lorentz boosts and proper spatial rotations. Such combinations always reduce to an equivalent single boost and single rotation,

$$\boldsymbol{\Lambda} = \mathbf{LR} = \begin{pmatrix} \gamma & -\boldsymbol{\beta}^T\gamma \\ -\boldsymbol{\beta}\gamma & \mathbf{L}_s \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_s \end{pmatrix} \tag{2.52}$$

where $\mathbf{L}_s$ and $\mathbf{R}_s$ are 3x3 matrices representing only the spatial components of these transformations (refer to (2.40) for the form of a general Lorentz boost). Note that the symmetry of these forms lead to the following equivalences,

$$\mathbf{L}^T\boldsymbol{\eta} = \begin{pmatrix} \gamma & -\boldsymbol{\beta}^T\gamma \\ -\boldsymbol{\beta}\gamma & \mathbf{L}_s \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -\mathbf{I} \end{pmatrix} = \begin{pmatrix} \gamma & \boldsymbol{\beta}^T\gamma \\ -\boldsymbol{\beta}\gamma & -\mathbf{L}_s \end{pmatrix} = \boldsymbol{\eta}\mathbf{L}^{-1} \tag{2.53a}$$

$$\mathbf{R}^T\boldsymbol{\eta} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_s^T \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -\mathbf{I} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & -\mathbf{R}_s^T \end{pmatrix} = \boldsymbol{\eta}\mathbf{R}^T = \boldsymbol{\eta}\mathbf{R}^{-1} \tag{2.53b}$$

Therefore, we have

$$\boldsymbol{\Lambda}^T\boldsymbol{\eta}\boldsymbol{\Lambda} = \mathbf{R}^T\mathbf{L}^T\boldsymbol{\eta}\mathbf{LR} = \mathbf{R}^T\boldsymbol{\eta}\mathbf{L}^{-1}\mathbf{LR} = \mathbf{R}^T\boldsymbol{\eta}\mathbf{R} = \boldsymbol{\eta}\mathbf{R}^{-1}\mathbf{R} = \boldsymbol{\eta} \tag{2.54}$$

Substituting this result into (2.51), we have

$$(c\tau')^2 = \mathbf{X}^T \left(\boldsymbol{\Lambda}^T\boldsymbol{\eta}\boldsymbol{\Lambda}\right)\mathbf{X} = \mathbf{X}^T\boldsymbol{\eta}\mathbf{X} = (c\tau)^2 \tag{2.55}$$

proving the desired invariance. In fact, the inner product or norm of all qualified four-vectors (those that transform via the Lorentz group) are similarly invariant. We will encounter several such four-vectors throughout this book.

If the proper time between two events — or rather, the square of the proper time interval, $(c\Delta\tau)^2$, between those two events — is negative, then it implies that the event occurs at a greater distance than can be traversed at the speed of light in the time allotted. Such events can have no causal relationship to one another. These intervals are called *spacelike separations*.

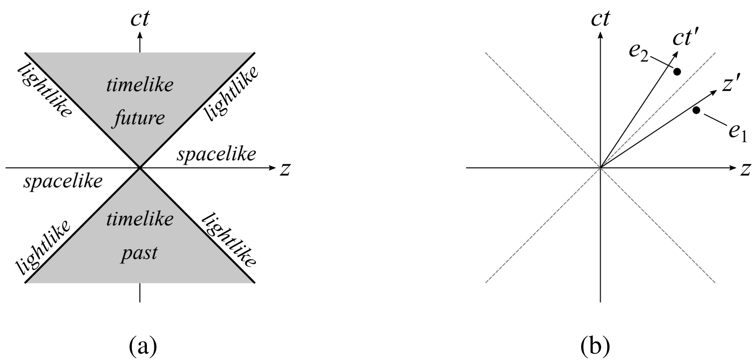


Figure 2.7 (a) Division of the Minkowski diagram into spacelike and timelike regions. (b) Illustration of the preservation of causality. An extreme Lorentz boost has managed to switch event e_1 from a future time coordinate in the rest frame to a past time coordinate in the moving frame, but this event is in the spacelike region and therefore cannot be causally related to an event at the origin. Event e_2, which is in the timelike region and may therefore have a causal relationship, cannot have its time coordinate inverted by any Lorentz boost.

However, if the proper time (squared) is positive, then it implies that more time separates them than is needed for light to traverse the space between. Such events may potentially have a causal relationship. These intervals are called *timelike* separations. The Minkowski diagram may therefore be separated into spacelike and timelike regions with respect to the origin, as shown in Figure 2.7(a). Events which lie on the boundary between spacelike and timelike events are called *lightlike*.

Note that only an event that lies in the spacelike region could potentially have its time coordinate change sign under a Lorentz boost transformation (as in Figure 2.7(b), event e_1), indicating that it would move from the past to the future, or vice versa, with respect to the origin. Since these events cannot be causally related, and since no transformation can change the value, much less the sign, of the proper time interval (effectively moving it from the spacelike region to the timelike region), causality is always preserved.

A closely related invariant number is the *spacetime interval*, s, found by negating the square norm,

$$s^2 = -(c\tau)^2 = x^2 + y^2 + z^2 - (ct)^2 \tag{2.56}$$

The spacetime interval would have been the positive norm of the four-position had we chosen a metric with the opposite signature (– + + +). Physically, either metric signature is valid.

2.3.6 Four-Velocity, Four-Momentum, and Energy

Consider the derivative of the four-position with respect to proper time,

$$\mathbf{U} = \frac{d\mathbf{X}}{d\tau} \tag{2.57}$$

The linearity property of the Lorentz transformation guarantees that the differential in the numerator is a four-vector that transforms in the same way as four-position, and since τ is a Lorentz invariant, the differential in the denominator is simply a constant, transformationally. Thus, the derived vector, $\mathbf{U}$, is also a qualified four-vector, in that it transforms in the same way, using the same Lorentz operator, $\mathbf{\Lambda}$, as four-position. This is called the *four-velocity*.

The constant $d\tau$ may be found as follows,

$$cd\tau = \sqrt{(cdt)^2 - dx^2 - dy^2 - dz^2} = dt\sqrt{c^2 - v_x^2 - v_y^2 - v_z^2} \tag{2.58a}$$

$$= dt\sqrt{c^2 - v^2} = \frac{cdt}{\cosh\zeta} \tag{2.58b}$$

Thus,

$$\mathbf{U} = \frac{d\mathbf{X}}{d\tau} = \frac{1}{d\tau}\begin{pmatrix} cdt \\ dx \\ dy \\ dz \end{pmatrix} = \frac{\cosh\zeta}{dt}\begin{pmatrix} cdt \\ dx \\ dy \\ dz \end{pmatrix} = \begin{pmatrix} c \\ v_x \\ v_y \\ v_z \end{pmatrix}\cosh\zeta \tag{2.59a}$$

$$= \begin{pmatrix} c \\ \mathbf{v} \end{pmatrix}\cosh\zeta \tag{2.59b}$$

Take note of the normalization,

$$\mathbf{U}\cdot\mathbf{U} = \mathbf{U}\boldsymbol{\eta}\mathbf{U} = \left(c^2 - v^2\right)\cosh^2\zeta = c^2\left(1 - \tanh^2\zeta\right)\cosh^2\zeta = c^2 \tag{2.60}$$

The four-velocity is the unit tangent to a worldline on the Minkowski diagram.

The product of a massive object's four-velocity and its rest mass (the mass it has in a reference frame for which it is not moving) is known as *four-momentum*,

$$\mathbf{P} = m\mathbf{U} = \begin{pmatrix} mc \\ m\mathbf{v} \end{pmatrix} \cosh\zeta = \begin{pmatrix} E/c \\ \mathbf{p} \end{pmatrix} \tag{2.61}$$

where $\mathbf{p} = p_x\mathbf{x} + p_y\mathbf{y} + p_z\mathbf{z}$ is the momentum three-vector (and reduces to the classical momentum, $p = mv$, when the Lorentz factor is small). Analysis based on the *principle of least action* [10, 14–16] shows that the first, temporal component[9] is associated with the system's total energy, E. The normalization of the four-momentum yields a general expression for the total energy of the system,

$$\mathbf{P} \cdot \mathbf{P} = m^2\mathbf{U} \cdot \mathbf{U} \tag{2.62a}$$

$$\left(\frac{E}{c}\right)^2 - p^2 = m^2c^2 \tag{2.62b}$$

$$\therefore E = \sqrt{p^2c^2 + m^2c^4} \tag{2.62c}$$

In the rest frame, where $p = 0$, we find that the energy is given by $E = mc^2$, otherwise known as the *rest energy*. This is, of course, Einstein's most famous equation establishing mass-energy equivalence [17].

Relativistically, the *four-force* on an object is equal to the rate of change of its four-momentum with respect to proper time,

$$\mathbf{F} = \frac{d\mathbf{P}}{d\tau} = m\frac{d\mathbf{U}}{d\tau} = m\mathbf{A} \tag{2.63}$$

where $\mathbf{A}$ is the *four-acceleration*. One must not neglect the time derivative of the Lorentz factor in this calculation,

$$\frac{d\cosh\zeta}{dt} = \frac{d}{dt}\frac{1}{\sqrt{1-\left(\frac{v}{c}\right)^2}} = \frac{\frac{d}{dt}\left(\frac{v}{c}\right)^2}{2\left(1-\left(\frac{v}{c}\right)^2\right)^{3/2}} \tag{2.64a}$$

$$= \frac{\frac{d}{dt}(\mathbf{v}\cdot\mathbf{v})}{2c^2\left(1-\left(\frac{v}{c}\right)^2\right)^{3/2}} = \frac{\mathbf{a}\cdot\mathbf{v}}{c^2}\cosh^3\zeta \tag{2.64b}$$

9 The elements of four-vectors are generally divided into temporal and spatial components, referring to the first entry and the last three, respectively, though their relationship to time and space may not always be easy to see.

Therefore,

$$\mathbf{A} = \frac{d\mathbf{U}}{d\tau} = \frac{d\mathbf{U}}{dt}\cosh\zeta = \begin{pmatrix} 0 \\ \mathbf{a} \end{pmatrix}\cosh^2\zeta + \begin{pmatrix} c \\ \mathbf{v} \end{pmatrix}(\cosh\zeta)\frac{d\cosh\zeta}{dt} \tag{2.65a}$$

$$= \begin{pmatrix} 0 \\ \mathbf{a} \end{pmatrix}\cosh^2\zeta + \begin{pmatrix} c \\ \mathbf{v} \end{pmatrix}\frac{\mathbf{a}\cdot\mathbf{v}}{c^2}\cosh^4\zeta \tag{2.65b}$$

where $\mathbf{v}$ and $\mathbf{a}$ are the classical velocity and acceleration three-vectors, respectively. Note that in an inertial reference frame matching the instantaneous velocity, such that $v = 0$ and $\cosh\zeta = 1$, we have

$$\mathbf{A} = \begin{pmatrix} 0 \\ \mathbf{a} \end{pmatrix} \tag{2.66}$$

in accord with the classical acceleration.

2.4 POINCARÉ'S COORDINATE TIME AND OTHER VARIANTS

It should be noted that an alternate formulation of the spacetime four-vector due to Poincaré has also been used in which the time component, called coordinate time, is imaginary,

$$\mathbf{X} = \begin{pmatrix} jct \\ x \\ y \\ z \end{pmatrix} \tag{2.67}$$

and the Lorentz boost transformation operator is likewise modified [5],

$$\mathbf{L} = \begin{pmatrix} \gamma & j\beta\gamma & 0 & 0 \\ -j\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.68}$$

This obviates the need for the Minkowski metric since the standard inner product of the four-vector automatically has the form given in (2.56). Nevertheless, the form presented in this chapter using all real numbers is the form that is most widely used today.

It is also true that some authors order the coordinates of their four-vectors differently, with the time coordinate last and the spatial coordinates first, while

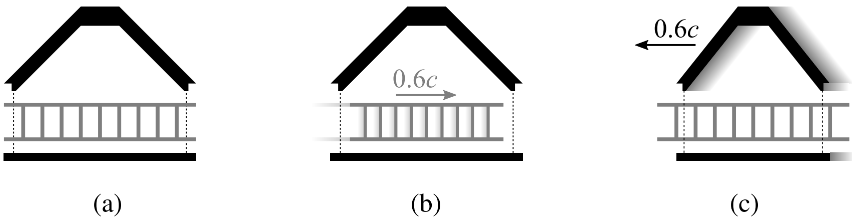


Figure 2.8 Illustration of the ladder paradox. (a) Ladder and barn at rest. The ladder is too long to fit between the doors, shown with dashed lines. (b) Length contraction when the ladder is moving at 60% the speed of light allows it to fit, and the barn doors to close, momentarily. (c) Length contraction when the barn is moving at 60% of the speed of light makes it less able to fit the ladder inside, nor can the doors both close simultaneously.

still others use different versions of the metric signature, for example, $(---+)$ or $(-+++)$. All such formulations are essentially equivalent for the purposes of this book, but I have chosen conventions that I think are the most useful and easiest to adapt to by readers who, like myself, have a background in classical electromagnetics. A summary of the effects that the choice of metric signature has on the key results of this book is given in Appendix E.

2.5 RESOLUTION OF APPARENT PARADOXES

Although my intent with this book is not to focus on the kinematic aspects of special relativity (as so many books already do), nor to defend it against the traditional attacks from those that have understandable difficulty accepting it, I think it is prudent to briefly discuss just two of the most famous apparent paradoxes that the theory raises. The resolutions of those paradoxes are instructive in themselves, and good practice in acclimating one's mind to the new way of thinking that special relativity requires.

2.5.1 The Ladder Paradox

The first apparent contradiction that we will discuss is known as the ladder paradox. We imagine having a ladder, say, 5 meters long, and a barn in which we wish to store it, but the barn has only 4.5 meters of storage space between the doors at the front and back of the building, as illustrated in Figure 2.8(a). However, knowing about Lorentz contraction, we believe that we can make the ladder fit, at least

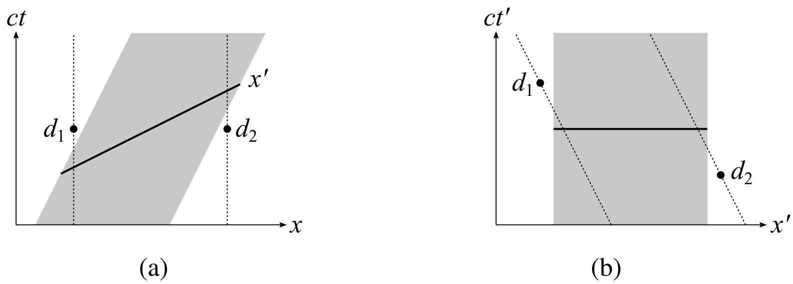


Figure 2.9 Minkowski diagrams explaining the resolution of the ladder paradox. Dashed lines indicate the barn doors, and the gray shaded region represents the ladder. Events d_1 and d_2 are the momentary closings of the first and second doors, respectively. (a) The rest frame of the barn. The two door closing events are simultaneous, and the ladder fits between them. The black line represents the contour of simultaneity for the ladder in its own moving reference frame. (b) The moving reference frame of the ladder. At the moment indicated by the black line, door 2 has already closed and reopened, while door 1 has not yet closed.

momentarily, by hurling the ladder at 60% the speed of light, or $\tanh\zeta = 0.6$. According to Lorentz, the ladder should now have a contracted length of

$$\Delta x = \frac{\Delta x'}{\cosh\zeta} = \Delta x'\sqrt{1 - \tanh^2\zeta} = (5\text{m})\sqrt{1 - 0.6^2} = 4\text{m} \tag{2.69}$$

which is short enough to fit in the barn, as shown in Figure 2.8(b). We may even imagine rigging the doors to both close simultaneously for a split second and then reopening again as the ladder passes through, emphasizing that the ladder was completely inside the barn for a brief period of time.

The source of the paradox is the principle that no inertial reference frame should have any special status. We may therefore analyze the situation from the alternate view of the ladder's reference frame, in which the barn itself is moving toward it at $0.6c$. In that reference frame, it is the barn that undergoes a length contraction, from 4.5m down to 3.6m, having even less capacity now to fit the 5m ladder. How, then, could both doors be shut at the same time?

To resolve this paradox, we must recognize that fitting the ladder into the barn means having both the front and the back ends inside the barn at the same time, or both doors shut at the same time, pairs of events that are separated in space and therefore cannot be said to be simultaneous in all reference frames. This is best illustrated by the Minkowski diagrams in Figure 2.9. The barn doors are once again indicated by dashed lines, while the ladder is shown with a gray shaded

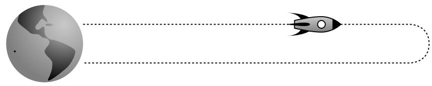


Figure 2.10 The twin paradox. One twin leaves Earth in a spaceship traveling at 60% the speed of light. After 10 years (as measured on Earth), he turns around and returns at the same speed. Which twin is older when they reunite?

box. In the reference frame of the barn, shown in Figure 2.9(a), the ladder passing through it is short enough to fit between the doors at a single moment in time. The door-shuttering events are denoted d_1 and d_2, and occur at the same vertical time coordinate, indicating simultaneity in this reference frame. However, the contour of simultaneity for the ladder, the black line labeled x', shows that, in the ladder's own reference frame, it extends beyond the horizontal position of both doors. The ladder's frame is shown more clearly in Figure 2.9(b), where it is evident that the two door-shuttering events are no longer simultaneous. Event d_2 has come and gone by the time the front-end of the ladder has cleared the building, while event d_1 has not yet occurred; the back door has shut and reopened again before the front door has even been shut.

2.5.2 The Twin Paradox

Perhaps even more well known to the layman is the twin paradox. In this scenario, we imagine two brothers, born as twins, one of whom embarks on a journey by spaceship while the other stays on Earth. The traveling twin moves away from Earth at 60% the speed of light for a period of 10 years (from the Earth's perspective), then turns around and returns at the same speed, shown in Figure 2.10. The Earth twin is now 20 years older, but observes his brother to have aged more slowly due to time dilation,

$$\Delta t = \Delta t' \sqrt{1 - \tanh^2 \zeta} = (20 \text{ yrs}) \sqrt{1 - 0.6^2} = 16 \text{ yrs} \tag{2.70}$$

The traveling twin is therefore younger than his brother, after the trip, by 4 years.

On the other hand, from the traveling twin's perspective, it is the Earth twin who moves away at the speed of $0.6c$, and then returns. In his reference frame, the Earth twin experiences time dilation, and should be 4 years younger when they get back together. Which twin is right? As is the case with most paradoxes, they are both right, at least in this case for the first half of the trip. On the outbound journey, the Earth twin would see the traveling twin age more slowly, while the traveling twin would see the Earth twin age more slowly.

Keep in mind that seeing in this context means to carefully make observations, correct for known factors such as time-of-flight, and then extrapolate what must have been the case after the fact. Both twins, for example, could be watching a streaming (but delayed) video feed of one another. The Earth twin would see his brother halfway through the trip as a younger version of himself engaging the thrusters on his spaceship to turn around. The Earth twin would not receive the image of that event until much later, but even correcting for the delay, the Earth twin would conclude that he had gained 2 years on his traveling brother by the time it had happened. From his perspective, the outbound leg took 10 years, but his traveling brother had only aged 8 years.

The traveling twin, also, would watch his brother aging more slowly on the video feed, and would know that the image on the screen when he turned his ship around was not live, but rather a delayed version of events. Nonetheless, upon waiting the appropriate amount of time for the video feed to propagate toward him, he would see that his brother did appear younger than himself when the course change had occurred. In fact, the traveling twin would experience an 8-year outbound journey, but would see his brother as having aged only $8 \times 0.8 = 6.4$ years during that time.

What neither of them has witnessed yet is the remarkable thing that happens when the traveling twin reverses course. At that moment, the traveling twin would observe (that is, he would eventually conclude from later observations and subsequent calculations) that his brother on Earth had rapidly aged 7.2 years in the short time it took him to turn around! He would then watch his brother age more slowly for the rest of the trip home, adding another 6.4 years in the apparently 8-year return trip. The traveling twin would experience 16 years round-trip, while his brother had aged $6.4 + 7.2 + 6.4 = 20$ years.

It is commonly said, by way of explanation, that the traveling twin's reference frame is not inertial, and thus not a valid reference frame under special relativity. The acceleration he experiences during the course change, however brief, is thus blamed for the discrepancy. To my way of thinking, this misses the greater point, which is that, in the context of special relativity, the traveling twin does not spend the entire time in one reference frame at all, but rather he switches reference frames halfway through the trip. One cannot equate the distant timing of events (the moment back on Earth when his course change took place) in two separate reference frames.

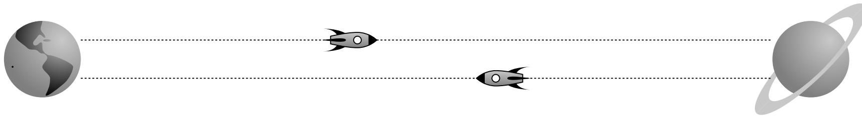


Figure 2.11 Variant of the twin paradox involving a third twin, or triplet.

To illustrate the point, I like to imagine a third twin (or triplet) traveling toward the Earth at the same speed from a planet twice as far away, as illustrated in Figure 2.11. There is no acceleration to point the finger at in this scenario, but there are three inertial reference frames: that of the Earth twin, that of the outbound twin, and that of the inbound twin. The two traveling twins may see each other through their portholes when they pass at the halfway point, and they would agree that they are the same age, but, most interestingly, they would not agree about the age of their brother back on Earth. The inbound twin would claim that their brother is 7.2 years older than the outbound twin thinks, and would feel vindicated when arriving back on Earth to see that he was right. The outbound twin, in contrast, would think his brother on Earth is 7.2 years younger than his inbound twin claims, but would additionally see the inbound twin aging even more slowly than their Earth brother, as a consequence of the combined speeds of their two ships, which would account for the difference in age of the Earth and inbound twins upon their reunion.

Nevertheless, there is actually no opportunity for the inbound twin and the outbound twin to compare notes. At the moment they pass, neither has yet been able to observe what is happening back on Earth; the video feed on which they are basing their eventual conclusions is still in transit. They would each complete their journey, having the evidence in their hands that they were correct about their conclusions, but unable to brag about it to their traveling brother.

A Minkowski diagram illustrating the passage of time from the Earth twin's perspective is shown in Figure 2.12(a). Dashed horizontal lines indicate the passage of years, 20 lines in total, from the traveling twin's departure until his return. The course change, from the Earth's perspective, occurs at the halfway point. In contrast, the contours of simultaneity from the traveling twin's perspective, marked with the slanted, dashed lines in Figure 2.12(b), show only 16 years passing between his departure and return. Importantly, the contours change slope discontinuously at the moment of the course change. Events back on Earth advance rapidly from 6.4 years into the journey to 13.6 years as he turns around. These diagrams are the same whether they show a single twin on the outbound and return legs of a two-way trip, or a pair of traveling twins crossing paths.

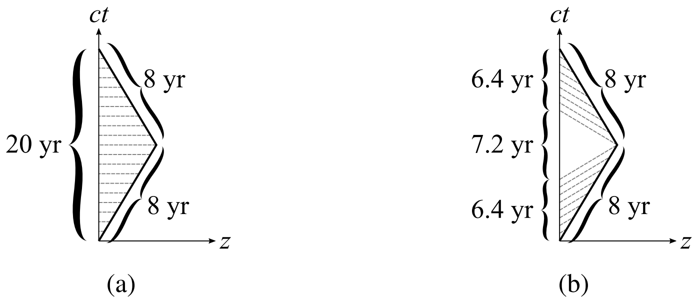


Figure 2.12 Minkowski diagram for the twin paradox. (a) Earth twin's contours of simultaneity marked out in single years, 20 in total. (b) Traveling twin's contours of simultaneity marked out in single years, 16 in total.

References

[1] B. R. Greene, *The Fabric of the Cosmos*. New York: Vintage Books, 2004.

[2] H. Genz, *Nothingness*. Massachusetts: Perseus Books, 1999.

[3] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge University Press, 1985.

[4] G. Arfken and H. Weber, *Mathematical Methods for Physicists*, 4th ed. San Diego, CA: Academic Press, 1995.

[5] M. Schwartz, *Principles of Electrodynamics*. New York: Dover, 1987.

[6] J. Schwinger, "A magnetic model of matter," *Science*, vol. 165, no. 3895, pp. 757–761, August 1969.

[7] Wikipedia. (2023, March) Fizeau experiment. https://en.wikipedia.org/wiki/Fizeau_experiment.

[8] R. Herman, "How fast is the earth moving?" *Scientific American*, October 1998. https://www.scientificamerican.com/article/how-fast-is-the-earth-mov/.

[9] Wikipedia. (2023, December) Michelson-Morley experiment. https://en.wikipedia.org/wiki/Michelson-Morley_experiment.

[10] L. Susskind and A. Friedman, *Special Relativity and Classical Field Theory: The Theoretical Minimum*. New York: Basic Books, 2017.

[11] Wikipedia. (2023, December) Lorentz transformation. https://en.wikipedia.org/wiki/Lorentz_transformation.

[12] Wikipedia. (2024, January) Hendrik Lorentz. https://en.wikipedia.org/wiki/Hendrik_Lorentz.

[13] I. M. Egdall, *Einstein Relatively Simple: Our Universe Revealed in Everyday Language*. Singapore: World Scientific, 2014.

[14] L. Susskind and G. Hrabovsky, *The Theoretical Minimum: What You Need to Know to Start Doing Physics*. New York: Basic Books, 2013.

[15] J. M. Charap, *Covariant Electrodynamics: A Concise Guide*. Maryland: Johns Hopkins University Press, 2011.

[16] C. Fox, *An Introduction to the Calculus of Variations*. New York: Dover Publications, 1987.

[17] I. Stewart, *In Pursuit of the Unknown: 17 Equations That Changed the World*. New York: Basic Books, 2012.

Chapter 3

Waves in Spacetime

Much of electrical engineering centers around the propagation of waves. It therefore seems prudent to consider how the frequency and wavelength of different kinds of waves transform from one reference frame to another. To do this, we must take into account not only how the wavelength changes under length contraction principles, but also how the propagation and reference frame velocities combine.

3.1 PARTIAL BOOSTS

Up to now, the examples we have explored involved switching reference frames between one in which an object was seen to be moving and another in which it is found at rest. Thus, the total boost velocity was the same as the initial object velocity. That is not always possible with waves, as some waves move at the speed of light (and others, like certain guided waves, may appear to move even faster [1–3]). One cannot catch up with such a wave. Instead, we must consider how the key properties of waves (e.g., wavelength and frequency) change when boosting between reference frames that only partially match the propagation of the wave. The frequency is affected by propagation velocity and time dilation, while wavelength is subject to length contraction.

3.1.1 Velocity Transformation

Due to the linearity of the Lorentz boost relationships, transformation of coordinate derivatives between reference frames is the same as the transformation of four-positions,

$$\frac{\partial \mathbf{X}'}{\partial t} = \mathbf{L}\frac{\partial \mathbf{X}}{\partial t} \tag{3.1}$$

Thus, for a boost in the x direction, we have

$$\frac{\partial ct'}{\partial t} = c\cosh\zeta - \frac{\partial x}{\partial t}\sinh\zeta = c\cosh\zeta - v_x\sinh\zeta \tag{3.2a}$$

$$\frac{\partial x'}{\partial t} = \frac{\partial x}{\partial t}\cosh\zeta - c\sinh\zeta = v_x\cosh\zeta - c\sinh\zeta \tag{3.2b}$$

$$\frac{\partial y'}{\partial t} = \frac{\partial y}{\partial t} = v_y \tag{3.2c}$$

$$\frac{\partial z'}{\partial t} = \frac{\partial z}{\partial t} = v_z \tag{3.2d}$$

where the rapidity parameterization has been used. However, these equations do not represent the apparent velocity in the transformed reference frame. To calculate that, we should have taken the derivative with respect to the transformed time coordinate, t', not the original time coordinate, t. We may derive the desired result by dividing (3.2b) to (3.2d) by (3.2a),

$$v'_x = \frac{\partial x'}{\partial t'} = \frac{\partial x'}{\partial t}\frac{\partial t}{\partial t'} = \frac{v_x\cosh\zeta - c\sinh\zeta}{c\cosh\zeta - v_x\sinh\zeta}c \tag{3.3a}$$

$$v'_y = \frac{\partial y'}{\partial t'} = \frac{\partial y'}{\partial t}\frac{\partial t}{\partial t'} = \frac{cv_y}{c\cosh\zeta - v_x\sinh\zeta} \tag{3.3b}$$

$$v'_z = \frac{\partial z'}{\partial t'} = \frac{\partial z'}{\partial t}\frac{\partial t}{\partial t'} = \frac{cv_z}{c\cosh\zeta - v_z\sinh\zeta} \tag{3.3c}$$

As we can see, apparent velocity does not transform linearly. Note that when $v_x = 0$, velocities in the orthogonal directions (y and z) are reduced by the Lorentz factor, $\cosh\zeta$. This can be understood as a time dilation effect; from the point of view of a moving observer, the progression of events outside his own reference frame occurs more slowly.

The x component of velocity — or, more specifically, the component of velocity in the same direction as the Lorentz boost — is in many ways the most interesting. Let us suppose that an object is moving at velocity v_1 with respect to the rest frame in the positive x direction ($v_x = v_1$), but that we are observing from an inertial reference frame that is moving at velocity v_2 in the opposite direction ($c\tanh\zeta = -v_2$). At what velocity do we see the object receding from us? In the Galilean world, we would simply add the two velocities, $v_1 + v_2$, but that would not be correct. According to (3.3a),

$$v'_x = \frac{v_x\cosh\zeta - c\sinh\zeta}{c\cosh\zeta - v_x\sinh\zeta}c = \frac{v_x - c\tanh\zeta}{c - v_x\tanh\zeta}c = \frac{v_1 + v_2}{1 + \frac{v_1 v_2}{c^2}} \tag{3.4}$$

This is called a *velocity addition formula*, and has a form which guarantees that two subluminal velocities cannot be combined in a way to produce a net velocity faster than light.

It takes on a much simpler form if we express both velocities in terms of rapidity,

$$v'_x = c\tanh\zeta' = c\frac{\tanh\zeta_1 + \tanh\zeta_2}{1 + \tanh\zeta_1\tanh\zeta_2} = c\tanh(\zeta_1 + \zeta_2) \tag{3.5}$$

Therefore, while the velocities do not add directly, the rapidities do — another bonus to using the rapidity formulation in performing complex calculations.

It was stated with no little fanfare in the first two chapters of this book that light was special, having the unique and unusual property of exhibiting constant propagation speed in all reference frames. This might give the reader the impression that light obeys one set of rules, while another set of rules applies to everything else. We may now see that this is an overstatement. Any object, no matter what its nature, would have the same constant-speed property as light does if it were accelerated up to a speed approaching c.

Example: Let us consider a mundane material object, say, a tennis ball, accelerated to a speed of $v_b = 0.999c$ (without concerning ourselves with the tremendous amount of energy it would take to accomplish it). If we transform to a new reference frame having speed $v = \pm 0.6c$, where the sign depends on whether the reference frame moves along with the tennis ball or in opposition to it, we find from (3.4) the range of possible speeds the tennis ball may attain in this new frame,

$$\frac{v_b - v}{1 - \frac{v_b v}{c^2}} \leq v'_b \leq \frac{v_b + v}{1 + \frac{v_b v}{c^2}} \tag{3.6a}$$

$$\frac{0.999c - 0.6c}{1 - 0.999 \cdot 0.6} \leq v'_b \leq \frac{0.999c + 0.6c}{1 + 0.999 \cdot 0.6} \tag{3.6b}$$

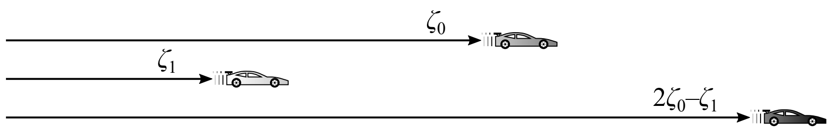


Figure 3.1 Illustration of relative rapidity in length contraction. The gray car, moving with rapidity ζ_0 is equally contracted in the reference frames of the white and black cars, whose relative rapidities are both $|\zeta_0 - \zeta_1|$.

$$0.996c \leq v_b' \leq 0.9997c \tag{3.6c}$$

As we can see, the ball's speed has been hardly affected by our tremendous boost in reference frame. The closer the ball's original speed was to the speed of light, the narrower the window of possible modified speeds becomes. We may thus conclude that light is simply a degenerate case under the common set of rules obeyed by all things, as it alone is capable of achieving the ultimate speed, c, at which no further boosts in velocity have any effect.

3.1.2 Partial Length Contraction

Length contraction is not merely a function of an object's length in a particular reference frame and the velocity of the boost, it depends also on the speed of the object in the original reference frame. Take, for example, the gray race car shown in Figure 3.1. It is moving with a rapidity of ζ_0 relative to the rest frame, in which we measure its length to be L. We might think that getting in the white race car, moving with rapidity ζ_1 with respect to the rest frame, would scale its length by a factor $\cosh \zeta_1$, but that is not the case. The fact that it was already moving in the original measurement matters. To find its true length in the reference frame of the white car, given an initial measurement in the rest frame, we must first expand its length by a factor $\cosh \zeta_0$ by first boosting into a comoving reference frame with the gray car, and then contract it again by $\cosh(\zeta_0 - \zeta_1)$, the Lorentz factor associated with the residual rapidity between the two cars. That is,

$$L' = \frac{L \cosh \zeta_0}{\cosh(\zeta_0 - \zeta_1)} = \frac{L \cosh \zeta_0}{\cosh \zeta_0 \cosh \zeta_1 - \sinh \zeta_0 \sinh \zeta_1} \tag{3.7a}$$

$$= \frac{L}{\cosh \zeta_1 (1 - \tanh \zeta_0 \tanh \zeta_1)} \tag{3.7b}$$

Figure 3.2 Illustration of relativistic aberration using the analogy of a car driving through the rain. (a) The rest frame, in which the rain falls vertically. (b) In the driver's reference frame, the rain appears to slant toward the car.

This formula is useful for determining the length contraction or expansion of an object that is already in motion in the base reference frame, as the gray car was in this case. Note that an identical result (in terms of length contraction) would be obtained by the black car, which surpasses the gray car's speed by the same amount that the white car fell short of it. In both cases, the difference in rapidity is $|\zeta_0 - \zeta_1|$.

3.1.3 Relativistic Aberration

Importantly, since the components of the velocity vectors do not all scale uniformly, a Lorentz boost that is not aligned with the direction of that velocity not only causes its magnitude to change, but alters its direction as well. This distortion effect is known as aberration. Aberration, like many of the effects I will describe in this chapter, is not exclusively a relativistic effect, but its exact form and magnitude are slightly altered by relativistic principles. Take, for example, Figure 3.2 showing a car driving through a rainstorm. In the rest frame, Figure 3.2(a), the rain is falling vertically, in what we will call the y direction. However, in the frame of the car moving rapidly in the x direction, the rain slants toward it, or so it appears to the driver.

More generally, let us write the velocity of the rainfall in the rest frame as

$$\mathbf{v} = v_x\mathbf{x} + v_y\mathbf{y} = v\left(\cos\theta_0\mathbf{x} + \sin\theta_0\mathbf{y}\right) \tag{3.8}$$

We may then calculate the velocity of the rainfall in the driver's reference frame, v', by applying (3.3) with the rapidity, ζ, of the car's movement in x. In particular, the cosine of the new velocity angle is

$$\cos\theta_1 = \frac{v'_x}{v'} = \frac{v_x\cosh\zeta - c\sinh\zeta}{c\cosh\zeta - v_x\sinh\zeta}\frac{c\cosh\zeta - v_x\sinh\zeta}{\sqrt{v_y^2 + \left(v_x\cosh\zeta - c\sinh\zeta\right)^2}} \tag{3.9a}$$

$$= \frac{v\cos\theta_0 - c\tanh\zeta}{\sqrt{v^2\sin^2\theta_0\operatorname{sech}^2\zeta + (v\cos\theta_0 - c\tanh\zeta)^2}} \tag{3.9b}$$

$$= \frac{v\cos\theta_0 - c\tanh\zeta}{\sqrt{(v^2-c^2)\operatorname{sech}^2\zeta + v^2\cos^2\theta_0\tanh\zeta^2 + c^2 - 2vc\cos\theta_0\tanh\zeta}} \tag{3.9c}$$

$$= \frac{v\cos\theta_0 - c\tanh\zeta}{\sqrt{(c - v\cos\theta_0\tanh\zeta)^2 - (c^2-v^2)\operatorname{sech}^2\zeta}} \tag{3.9d}$$

Note that if it was not rain falling on the car, but light, such that $v = c$, then this becomes

$$\cos\theta_1 = \frac{\cos\theta_0 - \tanh\zeta}{1 - \cos\theta_0\tanh\zeta} \tag{3.10}$$

or, as it is more commonly written [4],

$$\cos\theta_1 = \frac{\cos\theta_0 - \beta}{1 - \beta\cos\theta_0} \tag{3.11}$$

3.1.4 Four-Wavevector

With these principles in hand, we may now begin to understand how wave characteristics transform between reference frames. Suppose that we start in a meaningful rest frame in which a propagating wave has a known angular frequency, ω, and a phase velocity, v_p. We need not assume that the phase velocity is the speed of light (many guided waves do not propagate at the speed of light; some even seem to propagate faster!), nor that the wave is even electromagnetic in nature. Let us define the *wavenumber*, $k = 2\pi/\lambda$, where λ is the wavelength [1, 5]. This further implies that $\omega = kv_p$. To aid in our calculations, let ζ_p be the rapidity associated with the motion of the wave relative to the rest frame (i.e., $v_p = c\tanh\zeta_p$), and ζ_f is the rapidity of a moving reference frame we wish to convert to. According to (3.7), we may calculate the transformed wavelength as

$$\lambda' = \frac{\lambda}{\cosh\zeta_f\,(1 - \tanh\zeta_p\tanh\zeta_f)} = \frac{\lambda c}{c\cosh\zeta_f - v_p\sinh\zeta_f} \tag{3.12}$$

Therefore, the wavenumber transforms as

$$k' = k\cosh\zeta_f - \frac{kv_p}{c}\sinh\zeta_f = k\cosh\zeta_f - \frac{\omega}{c}\sinh\zeta_f \tag{3.13}$$

We can also calculate the new phase velocity using (3.3),

$$v'_p = \frac{v_p \cosh\zeta_f - c\sinh\zeta_f}{c\cosh\zeta_f - v_p\sinh\zeta_f}c = \frac{v_p\cosh\zeta_f - c\sinh\zeta_f}{\lambda/\lambda'} \tag{3.14a}$$

$$= (v_p\cosh\zeta_f - c\sinh\zeta_f)\,\frac{k}{k'} \tag{3.14b}$$

$$\therefore\ \frac{k'v'_p}{c} = \frac{\omega'}{c} = \frac{kv_p}{c}\cosh\zeta_f - k\sinh\zeta_f = \frac{\omega}{c}\cosh\zeta_f - k\sinh\zeta_f \tag{3.14c}$$

Thus,

$$\begin{pmatrix} \frac{\omega'}{c} \\ k' \end{pmatrix} = \begin{pmatrix} \cosh\zeta & -\sinh\zeta \\ -\sinh\zeta & \cosh\zeta \end{pmatrix} \begin{pmatrix} \frac{\omega}{c} \\ k \end{pmatrix} \tag{3.15}$$

We may recognize this as a Lorentz transformation, specifically for the case when the wave propagation, k, is aligned with the boost direction. We can generalize this result into a complete four-vector known as the *four-wavevector* that transforms according to the Lorentz equations,

$$\mathbf{K} = \begin{pmatrix} \frac{\omega}{c} \\ \mathbf{k} \end{pmatrix} = \begin{pmatrix} \frac{\omega}{c} \\ k_x \\ k_y \\ k_z \end{pmatrix} \tag{3.16}$$

Like all four-vectors, the elements make plain the duality between spatial and temporal effects; in this case, the frequency, ω, defines the temporal progression of phase in radians, while the wavenumber, $\mathbf{k}$, describes the spatial progression of phase. Note that the spatial and temporal components are related through the velocity as

$$\mathbf{k} = \frac{\omega}{v}\mathbf{n} = \frac{\omega}{v^2}\mathbf{v} \tag{3.17}$$

where $\mathbf{n}$ is the unit normal vector pointing in the direction of propagation, and $\mathbf{v}$ is the velocity. Therefore, the norm of the four-wavevector is

$$\mathbf{K}\cdot\mathbf{K} = \left(\frac{\omega}{c}\right)^2 - \mathbf{k}\cdot\mathbf{k} = \frac{\omega^2}{c^2} - \frac{\omega^2}{v^2} = \omega^2\frac{v^2 - c^2}{c^2v^2} = -\frac{\omega^2}{c^2\sinh^2\zeta} \tag{3.18}$$

We know that the final expression is a constant, because $\mathbf{K}$ is a qualified four-vector, whose norm must be invariant under Lorentz group transformations. We may therefore write

$$\mathbf{K} \cdot \mathbf{K} = -\frac{\omega^2}{c^2 \sinh^2 \zeta} = -\left(\frac{\omega_0}{c}\right)^2 \tag{3.19}$$

Lightlike waves, or waves traveling at speed c, have norm $\mathbf{K} \cdot \mathbf{K} = 0$. In that instance, the four-wavevector $\mathbf{K}$ is said to be *null*, or a *null vector*, despite having nonzero elements [6]. Waves traveling slower than light ($v < c$) have $\mathbf{K} \cdot \mathbf{K} < 0$ and are called spacelike or *slow waves*, while those traveling faster than light ($v > c$) have $\mathbf{K} \cdot \mathbf{K} > 0$ and are called timelike or *fast waves*.

The mere existence of timelike/fast waves may come as a surprise to some people, for the colloquial understanding is that nothing (no thing) can travel faster than light. Microwave engineers know to the contrary that many guided waves do have phase velocities that are faster than light, but this violates neither relativity nor causality because the wavefront in that case is nothing more than a superposition of plane waves, each traveling at a slower pace than their combined interference pattern. Energy and information in such waves travel at different speeds (e.g., the *group velocity* and *front velocity* [2]). Just as spectators collectively performing "the wave" in a stadium produce a wavefront that propagates faster than any individual person could move, the combined effect of multiple plane waves caroming off the sidewalls of a waveguide produce a mode that behaves as if it is traveling faster than light itself can move. It is evident from the equations in this section that the rapidity, ζ, and norm frequency, ω_0, for such timelike waves have values in the imaginary numbers.

3.2 DOPPLER EFFECTS

As anyone who has stood beside a highway and listened to cars speed past would know, the frequency of a wave is higher when the source is moving toward you than it is after the source has passed and is moving away from you. Known as the *Doppler effect*, this is a fundamental truth of all kinds of waves, whether it is the sound of a car engine, the light from a star, or the radio signal emitted by a wireless transmitter [7]. Relativistic analysis alters the magnitude of the effect, if not its qualitative behavior, when the velocity of either the source, the wave, or the observer is near the speed of light [8].

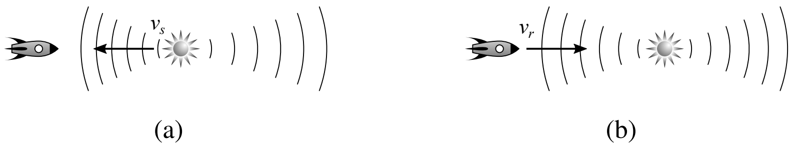


Figure 3.3 Illustration of the longitudinal Doppler effect. (a) The source, a star, moving toward the receiver, in a rocket, at speed v_s. (b) The receiver moving toward the source at speed v_r. In both cases, the phase velocity of the waves in the reference frame of the receiver is v_p.

3.2.1 Longitudinal Doppler Effect

We shall start with the simple case of a source that is moving directly toward a receiver at speed $v_s = c\tanh\zeta_s$, as shown in Figure 3.3(a). The phase velocity of the waves in the reference frame of the receiver is assumed to be v_p (it is not necessary in what follows that the wave is an electromagnetic one propagating at speed c). The frequency in the source's reference frame is f_s. Time dilation slows this oscillation down in the receiver's reference frame by the Lorentz factor corresponding to its motion, $\cosh\zeta_s$, so that the elapsed time between the emission of wavefronts in the receiver's reference frame is

$$\tau_r = \frac{\cosh\zeta_s}{f_s} \tag{3.20}$$

In that time interval, the preceding wavefront has advanced by a distance $v_p\tau_r$, while the source has advanced $v_s\tau_r = c\tau_r\tanh\zeta_s$. The new effective wavelength seen by the receiver is then given by the difference between these two,

$$\lambda_r = (v_p - v_s)\,\tau_r = \frac{v_p\cosh\zeta_s - c\sinh\zeta_s}{f_s} \tag{3.21}$$

and the new effective frequency is

$$f_r = \frac{v_p}{\lambda_r} = \frac{v_p}{v_p\cosh\zeta_s - c\sinh\zeta_s} f_s \tag{3.22}$$

Note that if $v_p = c$, then

$$\frac{f_r}{f_s} = \frac{1}{\cosh\zeta_s - \sinh\zeta_s} = \cosh\zeta_s + \sinh\zeta_s = \sqrt{\frac{1+\tanh\zeta_s}{1-\tanh\zeta_s}} \tag{3.23}$$

Alternatively, we could have used the four-wavevector to do this calculation. The wavevector in the receiver's reference frame is

$$\mathbf{K} = \begin{pmatrix} \frac{\omega_r}{c} \\ \mathbf{k}_r \end{pmatrix} = \begin{pmatrix} \frac{\omega_r}{c} \\ -\frac{\omega_r}{v_p} \\ 0 \\ 0 \end{pmatrix} \tag{3.24}$$

where the observed frequency is $\omega_r = 2\pi f_r$, and the x-component of the wavenumber, $\mathbf{k}_r$, is negative because the relevant wavefronts are propagating back toward the receiver (once the source has passed by the rocket, the sign of the wavenumber on the x axis will flip, and the following results would change). We may apply a Lorentz boost with rapidity $-\zeta_s$ to transform the wavevector into the source's reference frame,

$$\mathbf{K}' = \mathbf{L}\mathbf{K} = \begin{pmatrix} \cosh\zeta_s & \sinh\zeta_s & 0 & 0 \\ \sinh\zeta_s & \cosh\zeta_s & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\omega_r}{c} \\ -\frac{\omega_r}{v_p} \\ 0 \\ 0 \end{pmatrix} \tag{3.25a}$$

$$= \begin{pmatrix} \frac{\omega_r}{c}\cosh\zeta_s - \frac{\omega_r}{v_p}\sinh\zeta_s \\ \frac{\omega_r}{c}\sinh\zeta_s - \frac{\omega_r}{v_p}\cosh\zeta_s \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{\omega_s}{c} \\ \mathbf{k}_s \end{pmatrix} \tag{3.25b}$$

The source frequency may be found in the first element of the transformed wavevector above, ω_s/c. Therefore,

$$\frac{f_r}{f_s} = \frac{\omega_r/c}{\omega_s/c} = \frac{\omega_r/c}{\frac{\omega_r}{c}\cosh\zeta_s - \frac{\omega_r}{v_p}\sinh\zeta_s} = \frac{v_p}{v_p\cosh\zeta_s - c\sinh\zeta_s} \tag{3.26}$$

which is the same as (3.22).

If, instead, the receiver is moving toward the source at velocity v_r, as depicted in Figure 3.3(b), we may consider the wavevector originally in the source's reference frame,

$$\mathbf{K} = \begin{pmatrix} \frac{\omega_s}{c} \\ \mathbf{k}_s \end{pmatrix} = \begin{pmatrix} \frac{\omega_s}{c} \\ -\frac{\omega_s}{v_p} \\ 0 \\ 0 \end{pmatrix} \tag{3.27}$$

Note that this time we have assumed the velocity of the waves is v_p in the source frame, not the receiver frame. If the waves are light waves, then it does not matter, because $v_p = c$ in both frames. However, if the waves are something other than light, then this implies that the source is now in the rest frame of the medium.

Applying the Lorentz boost to transform into the receiver's reference frame,

$$\mathbf{K}' = \mathbf{L}\mathbf{K} = \begin{pmatrix} \cosh\zeta_r & -\sinh\zeta_r & 0 & 0 \\ -\sinh\zeta_r & \cosh\zeta_r & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\omega_s}{c} \\ -\frac{\omega_s}{v_p} \\ 0 \\ 0 \end{pmatrix} \tag{3.28a}$$

$$= \begin{pmatrix} \frac{\omega_s}{c}\cosh\zeta_r + \frac{\omega_s}{v_p}\sinh\zeta_r \\ -\frac{\omega_s}{c}\sinh\zeta_r - \frac{\omega_s}{v_p}\cosh\zeta_r \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{\omega_r}{c} \\ \mathbf{k}_r \end{pmatrix} \tag{3.28b}$$

Hence, the relative Doppler shift in this case may be written

$$\frac{f_r}{f_s} = \frac{\omega_r/c}{\omega_s/c} = \frac{1}{v_p}\left(v_p\cosh\zeta_r + c\sinh\zeta_r\right) \tag{3.29}$$

Note that when $v_p = c$ and $\zeta_r = \zeta_s$, this expression also becomes equivalent to (3.23), confirming the expectation under special relativity that the result should be independent of the reference frame. When $v_p \neq c$, the two expressions, (3.22) and (3.29), are not equivalent, because then this velocity is specified in relation to a medium, which is assumed at rest in either the receiver's or source's reference frames, respectively.

We may combine these two results into a more general expression for which both the source and the receiver are moving, at speeds $v_s = c\tanh\zeta_s$ and $v_r = c\tanh\zeta_r$, respectively. Conceptually, we can imagine first transforming from the moving source frame to that of an intermediate transceiver in the rest frame of the medium using (3.22), and then from that transceiver (or repeater) to the moving reference frame of the source using (3.29). However it is justified, the end result is just the product of those expressions,

$$\frac{f_r}{f_s} = \frac{v_p\cosh\zeta_r + c\sinh\zeta_r}{v_p\cosh\zeta_s - c\sinh\zeta_s} \tag{3.30}$$

It is instructive to factor this expression as follows,

$$\frac{f_r}{f_s} = \frac{v_p + c\tanh\zeta_r}{v_p - c\tanh\zeta_s}\cdot\frac{\cosh\zeta_r}{\cosh\zeta_s} = \left(\frac{v_p + v_r}{v_p - v_s}\right)\left(\frac{\cosh\zeta_r}{\cosh\zeta_s}\right) \tag{3.31}$$

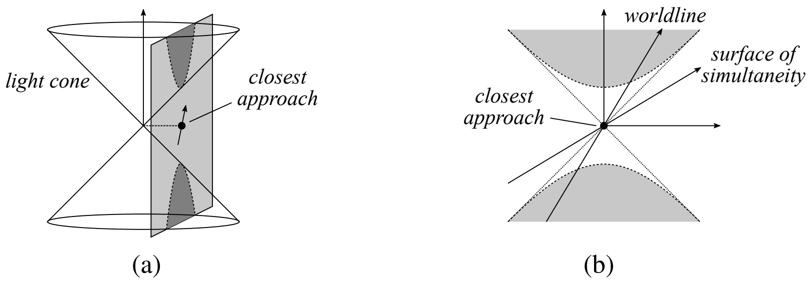


Figure 3.4 Minkowski diagrams for the transverse Doppler effect. (a) Three-dimensional diagram showing two spatial dimensions and one temporal dimension. Motion is assumed to occur entirely within the highlighted cutting plane. (b) Detail of the cutting plane, showing that the surface of simultaneity for the object in motion passes through the origin — thus, the point and moment of closest approach are identical in the reference frames of both the source and the receiver.

The first term in parentheses is the classical Doppler shift associated with convergent longitudinal motion [9] (motion for which positive velocities are measured when the source and receiver are moving toward one another), while the second term in parentheses is a relativistic correction factor due to time dilation. When the rapidities are small, the relativistic correction factor approaches unity. The longitudinal, relativistic Doppler shift is therefore only a small correction to the same basic phenomenon that is observed at nonrelativistic speeds.

3.2.2 Transverse Doppler Effect

In contrast, the classical version of the Doppler effect produces no frequency shift when the motion is transverse to the axis between source and receiver, but relativistic analysis predicts there is a shift, due at least in part to time dilation effects.

It may sound trivial, but to properly characterize this effect requires that we pay close attention to exactly what is meant by transverse. When a source and a receiver pass by one another in inertial reference frames, there comes a point at which they make their closest approach to one another, as shown in Figure 3.4. Since there is no motion along the line of sight when this condition is met, the moment is synchronous in both reference frames (unlike most events which cannot be said to be synchronous when spatially separated if there is any motion along the line of sight). Nevertheless, is the Doppler effect to be measured when they are actually at the closest approach, keeping in mind that the wave which is detected

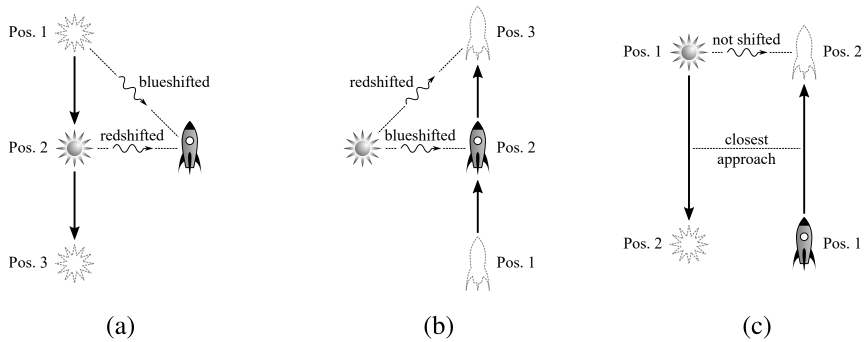


Figure 3.5 Illustrations of the transverse Doppler effect. (a) The source is in motion, while the receiver is stationary. (b) The receiver is in motion while the source is stationary. (c) Waves for which no shift is observed are neither emitted nor received at the closest point of approach.

at that moment must have been emitted earlier, or is it to be measured on the wave which is emitted from the point of closest approach, but detected later? The answer to that question will determine whether the wave is blueshifted or redshifted, from the receiver's perspective.

Consider Figure 3.5(a). We picture the source to be in motion, while the receiver is at rest. Waves emitted when the source was at position 1 arrive at the receiver when it reaches the point of closest approach, position 2. From the receiver's perspective, the motion of the source causes a time dilation that would tend to redshift the signal. However, that same motion has a forward component relative to the line of sight at position 1 that blueshifts it even more strongly, resulting in an overall blueshift. In contrast, the light emitted at the point of closest approach, position 2, has only the time dilation to affect it. From this position, the receiver measures a redshift. By the time it receives that redshifted signal, the source has already moved on to position 3.

For a light wave moving at speed c, the principle of relativity tells us that the same results should be obtained if we consider the receiver in motion instead of the source, as shown in Figure 3.5(b). When the receiver is at position 1, the source emits a wave that arrives when the two are at closest approach, position 2. In this case, time is dilated for the receiver, so that it notes a blueshift for this wave. At the same moment, more light is emitted, which reaches the receiver at position 3. The retreating motion of the receiver in this case overcomes the blueshift due to time dilation, resulting in an overall redshift.

In principle, we should be able to analyze the results in either reference frame, but the analysis is simplest when the object in motion, be it the source or the receiver, emits or receives the signal at its closest point of approach, so that no longitudinal motion is involved, only time dilation. Thus, for waves emitted from the closest point of approach (Figure 3.5(a) when the source is at position 2), time dilation at the source causes a redshift proportional to the Lorentz factor,

$$\frac{f_r}{f_s} = \frac{1}{\cosh \zeta} \qquad \text{for a wave emitted at closest point of approach} \tag{3.32}$$

On the other hand, waves that are received at the closest point of approach (Figure 3.5(b) when the receiver is at position 2), are blueshifted due to time dilation, and by the same factor,

$$\frac{f_r}{f_s} = \cosh \zeta \qquad \text{for a wave received at closest point of approach} \tag{3.33}$$

It is important to remember that (3.32) and (3.33) are for two different waves emitted/received at different times. Equation (3.32) describes a wave that was *emitted* at the closest point of approach, but received later, while (3.33) describes a wave that was received at the closest point of approach, having been emitted earlier.

We may use the same reasoning to deduce that no Doppler shift is observed for a transverse-directed wave when both the source and the receiver are in motion at the same velocity, if in opposite directions, as indicated in Figure 3.5(c). It should be noted that the waves in this case are neither emitted nor received at the point of closest approach, although the waves cross parallel to that line, offset in the direction from which the source came and toward which the receiver is headed.

3.2.3 Orbital Doppler Effects

These results can be used to analyze the scenario in which one source or receiver is in circular orbit around the other, as in Figure 3.6. Although these motions do not correspond to inertial reference frames, we can still apply the conclusions of special relativity to the point of emission or absorption, whichever is relevant, in an inertial reference frame that is tangential to the motion at that point in time. This is known as a *momentarily comoving reference frame* (MCRF).

When the source is in motion around the receiver, as in Figure 3.6(a), the situation is analogous to that in Figure 3.5(a) at position 2. The wave is redshifted by the Lorentz factor and the source has moved beyond its emission point by the

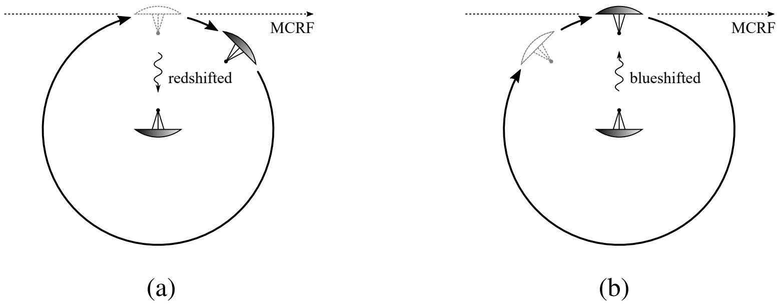


Figure 3.6 The Doppler effect for circular motions. (a) The source is in motion around the receiver, resulting in a redshift. (b) The receiver is in motion around the source, resulting in a blueshift. In each case, the momentarily comoving reference frame is shown with a dashed line.

time the wave reaches the receiver. In contrast, when the receiver is in motion around the source, as in Figure 3.6(b), the situation is analogous to Figure 3.5(b) at position 2. The wave is blueshifted at the moment the receiver catches up to it in its orbit.

Example: Earth's orbit around the Sun is much like the scenario depicted in Figure 3.6(b). The Sun appears blueshifted from our perspective by an amount commensurate with the orbital speed,[1]

$$\frac{f_r}{f_s} = \cosh\zeta = \frac{1}{\sqrt{1-\left(\frac{v}{c}\right)^2}} = \frac{1}{\sqrt{1-\left(\frac{67{,}000\text{ mph}}{299{,}792{,}458\text{ m/s}}\right)^2}} \tag{3.34a}$$

$$\approx 1.000000005 \tag{3.34b}$$

3.2.4 Doppler Effect with Arbitrary Linear Motions

We may generalize the above results to a situation involving arbitrary source and receiver trajectories and velocities as indicated in Figure 3.7. Let us assume that we are measuring these velocities relative to a reference frame where the speed of

1 That is to say, light from the Sun is blueshifted compared to how it would appear from Earth's distance if the Earth were not in motion. There is also a redshift associated with the light's emergence from the Sun's gravity well, and then a subsequent (but smaller) blueshift as it descends into Earth's gravity well, phenomena that can only be understood in the context of general relativity.

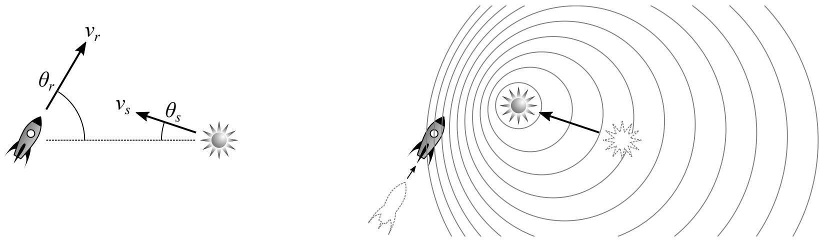


Figure 3.7 Setup for a general Doppler shift calculation. (a) Source and receiver trajectories. (b) Snapshot of wavefronts emitted by the source, showing also the displacement of both source and receiver during the time the waves were in transit. Although the diagram is drawn as if the trajectories of both source and receiver lie in the same plane (the plane of the paper), the calculation in the text does not depend on that being the case.

wave propagation in all directions is v_p; we will refer to this as the global reference frame. If the wave is something like a sound wave, then the global reference frame corresponds to the rest frame of the medium, but if it is a light wave in empty space, then any inertial reference frame meets this condition, where $v_p = c$.

Furthermore, for convenience, we will assume that the position of the source is drawn at its point of emission, while the receiver is shown at its point of reception — in reality, from the receiver's point of view, the source will have moved on by the time the wave is received, and the receiver will merely be catching up to the waves the source previously emitted.

Fortunately, the solution is quite simple once we understand the factors involved. The case for convergent longitudinal motion was already given in (3.31), where the first term corresponded to classical Doppler effects associated with the source and receiver moving relative to the propagation of the wavefronts, and the second term was a correction factor due to time dilation. All we have to do is replace the source and receiver velocities in the first term with their longitudinal projections, while leaving the rapidities in the second term unchanged since time dilation is independent of direction,

$$\frac{f_r}{f_s} = \left(\frac{v_p + v_r \cos\theta_r}{v_p - v_s \cos\theta_s}\right)\left(\frac{\cosh\zeta_r}{\cosh\zeta_s}\right) \tag{3.35}$$

Note that it is unnecessary that both velocities, $\mathbf{v}_r$ and $\mathbf{v}_s$, lie in the same plane. One should also keep in mind that the angles, θ_r and θ_s, are specific to the reference frame used for analysis; that is, the global reference frame or rest frame of the

Table 3.1

Transformation of Propagating Waves Between Reference Frames

Effect	Expression
Aberration, general	$\cos\theta_s' = \dfrac{v_p\cos\theta_s - c\tanh\zeta_r}{\sqrt{(c - v_p\cos\theta_s\tanh\zeta_r)^2 - \left(c^2 - v_p^2\right)\operatorname{sech}^2\zeta_r}}$
Aberration of light $v_p = c$	$\cos\theta_s' = \dfrac{\cos\theta_s - \tanh\zeta_r}{1 - \cos\theta_s\tanh\zeta_r} = \dfrac{\cos\theta_s - \beta_r}{1 - \cos\theta_s\beta_r}$
Doppler shift, general	$\dfrac{f_r}{f_s} = \left(\dfrac{v_p + v_r\cos\theta_r}{v_p - v_s\cos\theta_s}\right)\left(\dfrac{\cosh\zeta_r}{\cosh\zeta_s}\right)$
Doppler shift of light $v_p = c$	$\dfrac{f_r}{f_s} = \dfrac{\cosh\zeta_r + \cos\theta_r\sinh\zeta_r}{\cosh\zeta_s - \cos\theta_s\sinh\zeta_s}$
Doppler shift, longitudinal $\theta_r = \theta_s = 0$	$\dfrac{f_r}{f_s} = \left(\dfrac{v_p + v_r}{v_p - v_s}\right)\left(\dfrac{\cosh\zeta_r}{\cosh\zeta_s}\right)$
Doppler shift, transverse $v_s = 0, \theta_r = \pm\frac{\pi}{2}$	$\dfrac{f_r}{f_s} = \cosh\zeta_r$
Doppler shift, transverse $v_r = 0, \theta_s = \pm\frac{\pi}{2}$	$\dfrac{f_r}{f_s} = \dfrac{1}{\cosh\zeta_s}$

$v_p =$ phase velocity of the wave in the rest frame of the medium (if applicable).
$v_s, \zeta_s, \theta_s =$ velocity, rapidity, and projected angle of the source's motion.
$v_r, \zeta_r, \theta_r =$ velocity, rapidity, and projected angle of the receiver's motion.
$\beta_r = v_r/c$.

medium. The angle of approach seen by either the source or receiver in their own reference frames will be different as a consequence of aberration effects, described in Section 3.1.3.

A summary of the alterations that may apply to a propagating wave under reference frame transformation is given in Table 3.1.

3.3 GLOBAL NAVIGATION SATELLITE SYSTEMS

Global Navigation Satellite Systems (GNSS) are a rare example of current, real-world applications in which relativity plays a significant role [10–14]. Although special relativity alone is not sufficient to describe the operation of these systems

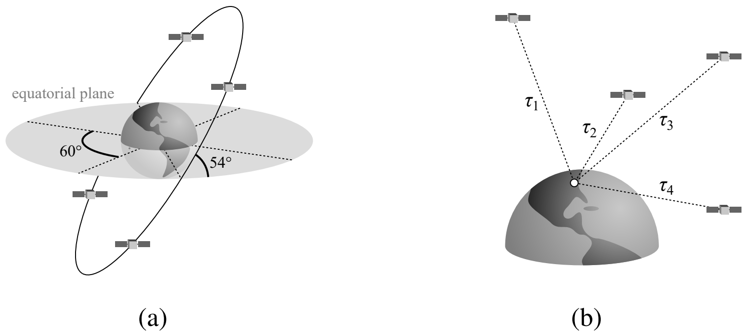


Figure 3.8 Overview of the Global Positioning System. (a) Typical satellite orbit, inclined 54 degrees with respect to the equatorial plane. There are six such orbits in total at 60-degree intervals, having intersections with the equatorial plane shown by the dashed lines. (b) Measurement of time delays from four satellites are sufficient to locate the receiver in spacetime.

(for general relativity is crucial as well) it is nevertheless an excellent illustration of the principles we have been studying in this chapter.

3.3.1 System Description

There are currently four global navigation satellite systems in operation. The Global Positioning System (GPS) operated by the United States Space Force [15] and the Global Navigation Satellite System (GLONASS) operated by the Russian Federation were the first to become operational. These were followed by China's BeiDou Navigation Satellite System, and finally Galileo operated by the European Union. In addition, there are a number of regional augmentation systems in place in some countries to enhance and extend global positioning capabilities in these areas [10]. For the purposes of this section, we shall focus on GPS.

The GPS satellite network consists nominally of 24 satellites, four each in six orbits, where each orbit is inclined 54 degrees with respect to the equatorial plane and distributed at 60-degree intervals azimuthally, as illustrated for one such orbit in Figure 3.8(a). They orbit at an altitude of 21,183 km, giving them an orbital period of very nearly 12 sidereal hours. This ensures that between five and eight satellites are visible from any point on Earth at all times, and that the satellites repeat their configuration roughly once per day (24 hours minus 4 minutes) [12]. In practice, new satellites are launched periodically, enhancing capability while replacing older ones that have failed. As of this writing, there are 38 usable satellites in orbit, 32 of which are in active use [15].

Each satellite carries four high-precision atomic clocks, two based on the Cesium standard and two on Rubidium. The accuracy of these clocks improves with each generation, roughly a factor of 10 every 10 years. The first Block I satellites had a relative rate error of about 1 part in 10^{12}. The current generation is closer to 5 parts in 10^{14}, corresponding to about 4 nanoseconds per day (30 cm in range at the speed of light) [11, 12]. Thus, any relativistic effects measured on a similar scale may be considered significant, and should be accounted for.

3.3.2 Multilateration

The basic function of the GPS system is *multilateration*, or the determination of coordinates given range measurements to some number of known points, as illustrated in Figure 3.8(b). For example, if the distances from a fixed point on the ground to three different satellites is known precisely (and the positions of those satellites is also known in some frame of reference), then the position coordinates of that fixed point in three-dimensional space can be calculated. To see this, one may imagine drawing a sphere around each satellite with radius equal to the distance measured from that satellite. The intersection of two of those spheres defines a circle, and the intersection of that circle with the third sphere yields, in general, only two possible points for the unknown location. One of those points is far out in space beyond the orbit of the satellites and may safely be discarded [13].[2]

However, the range in this case is measured by propagation delay, and accurate determination of those delays would require that the user on the ground is carrying an atomic clock with precision comparable to those on the satellites in space. Of course, this is entirely impractical. Instead, a fourth satellite is used to add one more measurement with which the user's clock can also be corrected. In effect, we have written down four equations to solve for four unknowns, namely the spacetime coordinates, (ct, x, y, z). We thus determine not only the user's position in space, but his or her four-position in spacetime, with atomic-clock accuracy. This makes satellite navigation systems an excellent tool not only for geopositioning, but for time transfer as well [12, 13].

2 In reality, the true range to each satellite is not known, rather the time-difference of arrival (TDOA) between a pair of satellites is known. The locus of points corresponding to a fixed TDOA is a hyperboloid of revolution rather than a sphere [15], but the principle is the same.

3.3.3 Relativistic Effects

The actual calculations involved in turning raw GPS delay measurements into geographical coordinates are quite complex. Setting relativity aside for the moment, one must account for the motion of both the satellites and the Earth during the time it takes for a signal to propagate from transmitter to receiver, atmospheric and ionospheric effects, perturbation of the satellites' orbits due to deviation of the Earth's shape (and its associated gravitational field) from that of a perfect sphere, and even the motion of the user if traveling in a vehicle, to name a few. My intention is not to go through these calculations in detail, rather I simply wish to quantify the relativistic effects in order to illustrate how they impact the precision of a real-world application.

Example: Consider that a user on the ground will see a transverse Doppler effect due to time dilation, causing a redshift, or slowing of the clock on board the satellite as a consequence of its rapid movement in orbit. The situation is like that illustrated in Figure 3.6(a). The magnitude may be calculated using (3.32)

$$\frac{f_r}{f_s} = \frac{1}{\cosh\zeta} = \sqrt{1 - \left(\frac{v_s}{c}\right)^2} \tag{3.36}$$

where $v_s = 3,874$ m/s is the typical orbital velocity of a GPS satellite. Before we plug in some numbers, however, it is useful to expand this into a Taylor series,

$$\frac{f_r}{f_s} \approx 1 - \frac{1}{2}\left(\frac{v_s}{c}\right)^2 - \frac{1}{8}\left(\frac{v_s}{c}\right)^4 + \cdots \tag{3.37}$$

when $v_s \ll c$. Note that in a purely Newtonian world, there would be no transverse Doppler effect, and we would retain only the constant term (=1) above. The second-order term is purely relativistic, and is a sufficient correction given the accuracy of our current atomic clocks. The relativistic Doppler correction is thus often referred to as the second-order Doppler correction for the purposes of satellite navigation,

$$\frac{\Delta f_r}{f_s} \approx -\frac{1}{2}\left(\frac{v_s}{c}\right)^2 = -\frac{1}{2}\left(\frac{3,874 \text{ m/s}}{299,792,458 \text{ m/s}}\right)^2 \approx -8.35 \times 10^{-11} \tag{3.38}$$

Recall that the stability of our on-board clocks was now approximately 5 parts in 10^{14}; this is thousands of times worse. Were we to neglect this term, the time received from GPS satellites would suffer a cumulative error of about 7 microseconds per day, equivalent to over 2 km per day error in range.

It must be noted that other important relativistic effects come into play that we have not detailed here, including, especially, gravitational effects on the clock speed, orbital mechanics of the spacecraft, and even curvature of the signal propagation path due to gravity. These effects are the subject of general relativity and, as such, a detailed explanation of them is beyond the scope of this book. However, it can be simply said that while clocks in motion run slower due to time dilation, those at high altitude run faster from the point of view of an observer on the ground. This blueshift is a consequence of the signal's descent into Earth's gravity well, and works against the second-order Doppler correction described above. In fact, the gravitational blueshift is dominant over the time-dilation redshift at the orbital distance of the GPS satellites. For spacecraft in low Earth orbit (like the Space Shuttle), where the orbital speed is high and the difference in gravitational potential is smaller, the time-dilation effect is dominant. As a point of interest, the two effects roughly cancel each other at an altitude a little over 3,000 km [16].

It is historically interesting that at the time the first Cesium atomic clock was launched into orbit (on June 23, 1977, many decades after Einstein published his special and general theories of relativity), there was still lingering debate and uncertainty regarding the magnitude of these corrections, their sign, and even whether relativity was a true effect that needed correcting in the first place [16]! The satellite was thus built with a frequency synthesizer capable of post-launch adjustment to account for the uncertainty. After 20 days in orbit, during which time the clock speed was carefully monitored, the synthesizer was turned on to put in the exact correction that relativity had predicted. In later generations, the synthesizer was omitted, and instead a fixed factory offset was put into the clocks on the ground before launch, anticipating the time dilation and gravitational blueshift that would occur once that satellite is put into orbit.

3.3.4 The Post-Newtonian Model

As the foregoing discussion suggests, the general approach of the GPS system is to assume, initially, a Newtonian geometric model with a fixed speed of light, c, and then account for relativistic effects via post-Newtonian (PN) corrections, like the second-order Doppler term above. This is likened to a first-order approximation of relativity, accurate when the motions of the transmitter and receiver are slow compared to the speed of light, and when the difference in gravitational potential between them is small [11].

Indeed, there is much about the conceptual framework of GPS that seems almost defiant in view of the fundamental lessons of relativity, such as the abolishment of global simultaneity. Consider, for example, that the GPS post-Newtonian model uses what is called an Earth-centered Earth-fixed (ECEF) reference frame for spatial coordinates, where fixed points in space correspond to a location on (and rotating with) the surface of the Earth, while the reference for temporal measurements is Earth-centered and inertial (ECI), fixed in orientation to a background of measured star locations. That is, the time frame is inertial, while the spatial frame is not [12]. Transference between these frames accounts for some of the complication inherent in GPS calculations. As one author writes [14], "is there anything less relativistic than the obstination to keep synchronized a system of clocks in relative movement?"

The best ground-based clocks now achieve stability close to 1 part in 3×10^{17} [11]. As technology marches forward, we can expect optical clocks of the future to achieve even higher standards of accuracy. This makes it feasible that global navigation satellite systems could, in principle, achieve positional accuracy of millimeters or even better, and exciting new as-yet-unimagined applications are the inevitable result. But persistence with the post-Newtonian model means that more and more complex corrections for subtle, higher-order relativistic effects will be needed in order to take advantage of them. This has led some researchers to propose a new paradigm that embraces our relativistic reality from the outset, so that no such corrective actions are needed. The system would be inherently exact and true (within the accuracy of the clocks themselves) for as long as relativity itself is deemed to be correct [12, 14].

3.4 DISPERSION IN MINKOWSKI SPACE

The analyses of boosts and Doppler effects in the first part of this chapter were sufficiently general to apply to waves of all kinds, whatever their physical basis (e.g., light, sound, water), and whatever their speed of propagation, even those that travel faster than c (the aforementioned fast waves). While it is typical in relativistic physics to talk only about light, and assume that all light travels at speed c, there are many situations in engineering in which this is not the case (consider the dominant mode of a rectangular waveguide, or the Gouy phase of a Gaussian beam [1, 3]). It is useful to familiarize ourselves with some techniques for visualizing the propagation behavior of all kinds of waves and how they are affected by Lorentz transformations.

3.4.1 Dispersion Diagrams

Consider the four-wavevector for two x-directed waves, one fast and one slow, with the wavenumber plotted on the horizontal axis and the radial frequency on the

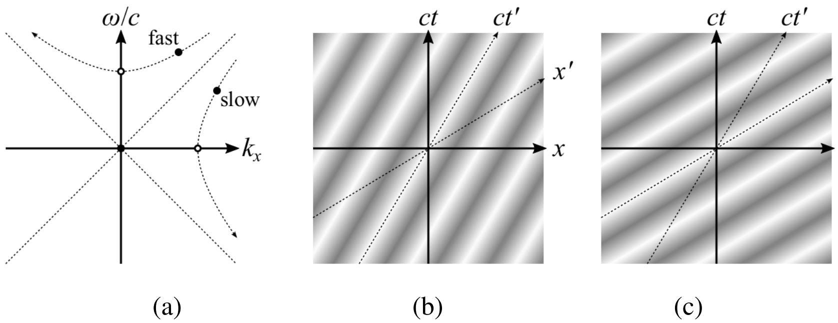


Figure 3.9 (a) Dispersion diagram showing boost loci for fast and slow waves. (b) Lorentz-invariant phase for a slow wave plotted as levels of gray (periodically, from white to gray over π radians and back again) on a conventional Minkowski diagram. The boost axes, x' and ct', correspond to a reference frame that has caught up to the wave. (c) Lorentz-invariant phase for a fast wave. The boost axes correspond to a reference frame that has expanded the instantaneous wavelength to infinity.

vertical axis, shown in Figure 3.9(a). Known in engineering circles as a *dispersion diagram*, this plot bears a strong resemblance to the Minkowski diagrams we studied in Chapter 2, only the units have been inverted (both k and ω/c have units of m^{-1}). Lightlike waves traveling at speed c would lie on the dashed, diagonal lines. The lower black dot closer to the k_x axis represents the four-wavevector of our slow wave. Boosts in the positive x direction would move this dot along the hyperbolic path shown. A boost of sufficient rapidity would catch up to the wave, leaving the four-wavevector with no temporal frequency component (represented by an open circle on the plot). Such a wave would appear frozen to the observer. A faster boost would see the wave receding from the observer's perspective, thus propagating (apparently) in the opposite direction.

Fast waves are represented in the diagram by points closer to the ω/c axis, such as the upper black dot in Figure 3.9(a). Lorentz boosts in the positive x direction move it leftward, as shown by the upper hyperbolic path, until a point is reached where its wavenumber, k_x, has reduced to zero. In effect, the wavelength has become infinitely long, from the observer's perspective. It still has a temporal frequency, however. Such a wave would appear constant throughout space while oscillating in value over time (like a horizontal line moving up and down as opposed to a sinusoidal curve moving left or right). Faster boosts would restore the sinusoidal form as the wavelength decreases from infinity, with the wave once again moving in the reverse direction.

Note that no boost is capable of converting a fast wave to a slow wave, or vice-versa. These qualitative characterizations are invariant for all kinds of waves.

3.4.2 Lorentz-Invariant Phase

The inner product of the four-wavevector, $\mathbf{K}$, and the four-position, $\mathbf{X}$, is called the Lorentz-invariant scalar phase of the wave, Φ,

$$\mathbf{K} \cdot \mathbf{X} = \omega t - \mathbf{k} \cdot \mathbf{r} = \Phi \tag{3.39}$$

One could thus write the equation for a wave of amplitude A as

$$A \cos(\omega t - \mathbf{k} \cdot \mathbf{r}) = A \cos(\mathbf{K} \cdot \mathbf{X}) \tag{3.40}$$

While $\mathbf{K}$ and $\mathbf{X}$ both transform identically under Lorentz boosts, the phase Φ remains constant, as evidenced by the symmetry properties of Lorentz transformations,

$$\mathbf{K}' \cdot \mathbf{X}' = (\mathbf{K}')^T \boldsymbol{\eta} \mathbf{X}' = (\mathbf{L}\mathbf{K})^T \boldsymbol{\eta} (\mathbf{L}\mathbf{X}) = \mathbf{K}^T \mathbf{L}^T \boldsymbol{\eta} \mathbf{L} \mathbf{X} \tag{3.41a}$$

$$= \mathbf{K}^T \boldsymbol{\eta} \mathbf{L}^{-1} \mathbf{L} \mathbf{X} = \mathbf{K}^T \boldsymbol{\eta} \mathbf{X} = \mathbf{K} \cdot \mathbf{X} \tag{3.41b}$$

The Lorentz-invariant phase can be plotted in grayscale on a conventional Minkowski diagram, as shown for a slow wave in Figure 3.9(b). An observer at $x = 0$ sees the wave at his location oscillate in time (the ct axis, or worldline, cuts through multiple fringes of phase) as well as in space (the x axis, or surface of simultaneity, also passes through fringes). A boost of sufficient velocity catches up to the wave, such that the worldline, ct', is now aligned with a phase fringe. There is no longer any variation in time, although the wave continues to have peaks and valleys in space (along x'). The wave appears frozen in time from this moving observer's perspective. The spatial wavelength is longer than before, but still finite.

A similar Minkowski diagram for fast waves is shown in Figure 3.9(c). In this case, boosts in the direction of propagation cause the waves to oscillate more slowly in time while wavelength increases until the latter becomes infinite. Thus, the x' axis at this point is aligned with a phase fringe, while the worldline ct' continues to pass through them. The wave has flattened out, from the observer's perspective, but continues to oscillate up and down.

In Chapter 4, we will begin to apply the consequences of the Lorentz transformation more directly to electromagnetic phenomenon. We will show how the electromagnetic field quantities themselves (e.g., $\mathbf{E}$ and $\mathbf{H}$) may be transformed from one inertial reference frame to another, and how to write Maxwell's equations in a form that is independent of reference frame.

References

[1] M. A. Morgan, *Principles of RF and Microwave Design*. Norwood, MA: Artech House, November 2019.

[2] P. W. Milonni, *Fast Light, Slow Light and Left-Handed Light*. New York: Taylor & Francis, 2005.

[3] P. F. Goldsmith, *Quasioptical Systems*. New York: IEEE Press, 1998.

[4] Wikipedia. (2024, January) Relativistic aberration. https://en.wikipedia.org/wiki/Relativistic_aberration.

[5] D. M. Pozar, *Microwave Engineering*, 4th ed. New York: Wiley, 2011.

[6] Wikipedia. (2023, November) Four-vector. https://en.wikipedia.org/wiki/Four-vector.

[7] S. Winchester, *The Perfectionists: How Precision Engineers Created the Modern World*. New York: Harper Collins, 2018.

[8] Wikipedia. (2024, January) Relativistic Doppler effect. https://en.wikipedia.org/wiki/Relativistic_Doppler_effect.

[9] Wikipedia. (2024, January) Doppler effect. https://en.wikipedia.org/wiki/Doppler_effect.

[10] Wikipedia. (2024, January) Satellite navigation. https://en.wikipedia.org/wiki/Satellite_navigation.

[11] Ž. Hećimović, "Relativistic effects on satellite navigation," *Technical Gazette*, vol. 20, no. 1, pp. 195–203, February 2013.

[12] J. F. Pascual-Sánchez, "Introducing relativity in global navigation satellite systems," *Annalen der Physik*, vol. 16, no. 4, pp. 258–273, April 2007.

[13] M. Pössel. (2023, August) Relativity and satellite navigation. https://www.einstein-online.info/en/spotlight/relativity-satellite-navigation/.

[14] A. Tarantola, L. Klimeš, J. M. Pozo, and B. Coll, "Introducing relativity in global navigation satellite systems," *Annalen der Physik*, pp. 5–35, October 2018.

[15] Wikipedia. (2024, January) Global positioning system. https://en.wikipedia.org/wiki/Global_Positioning_System.

[16] N. Ashby, "Relativity in the global positioning system," *Living Reviews in Relativity*, vol. 6, no. 1, pp. 1–42, January 2003.

Chapter 4

Covariant Electrodynamics

In Chapter 2, we introduced the concept of a four-vector for coordinates in space-time and their derivatives, and showed how these groups of variables can be transformed from one inertial reference frame to another using the Lorentz equations. In Chapter 3, we talked about another kind of four-vector, the four-wavevector, that describes the propagation properties of all kinds of waves and transforms according to the same rules. Now we will continue to build upon that foundation by defining other four-vector quantities pertaining specifically to electromagnetic phenomenon. We will also introduce a new type of object, the tensor, which may have additional degrees of freedom and transform in a different way, but based upon the same operator. Finally, these will then be used to rewrite the laws of electrodynamics in a form that is independent of inertial reference frame.

4.1 KINEMATICS OF MOVING CHARGES

Let us begin by thinking about charges and charges in motion (otherwise known as current). The relationship that these two quantities have to one another relativistically can be explored in the context of Lorentz contractions and time dilation.

4.1.1 Charge Density Transformation

Consider a stationary lump of charge, having charge density ρ, distributed uniformly throughout a cubic volume, as shown in Figure 4.1(a). We may think of this as a concentration of millions of elemental charges, for example, electrons. Imagine then that we convert to another reference frame where the charge is seen to be in

Figure 4.1 (a) A cubic lump of charge density, ρ, at rest. (b) The lump of charge in motion, having a new density, $\rho\gamma = \rho \cosh\zeta$, due to Lorentz contraction of its spatial distribution.

motion, as in Figure 4.1(b). Clearly, the moving charge now constitutes a current — even Galileo, had he understood the nature of electric current, would have predicted that — but, in addition, we see the charge density increase. The number of electrons has not changed, but length contraction along the longitudinal axis has compressed them into a smaller volume. The amount that the charge density has increased is simply the Lorentz factor, $\cosh\zeta$, or

$$\rho' = \rho \cosh\zeta \tag{4.1}$$

The current density may be written as the product of this new charge density times the velocity,

$$J'_x = -(\rho\cosh\zeta)v = -c\rho\cosh\zeta\tanh\zeta = -c\rho\sinh\zeta \tag{4.2}$$

where we have assumed that motion of the reference frame is in the $+x$ direction (so that the charge moves in the $-x$ direction).

4.1.2 Current Density Transformation

What if, instead of starting with a lump of charge at rest, we started with a neutral current (having net-zero charge density)? Since current is charge in motion, to make it neutral, we must balance that moving charge with an equal density of stationary charge having the opposite sign, as shown in Figure 4.2(a). (We could alternately have assumed that both sets of charge were in motion in opposite directions at half the speed, but the end result would be the same.) Upon transformation to the new reference frame, the negative charges are found to be at rest, while the positive charges are in motion. More importantly, the positive charge density has increased, because the spacing between them is Lorentz-contracted, while the negative charge density has decreased for essentially the same reason (that is, it was contracted in the first place, and is now expanded back to the rest spacing). The result is that,

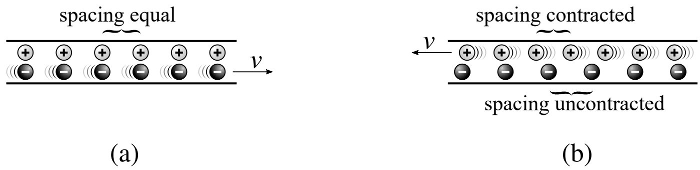


Figure 4.2 (a) A neutral current comprising positive charges at rest (the light-colored balls) and negative charges in motion (the dark balls). (b) After transformation to a new reference frame, the negative charges are at rest, and the positive charges are in motion. Additionally, Lorentz contraction reduces the spacing between positive charges, while the spacing between negative charges is expanded (uncontracted). This results in a net positive charge density.

although we started with no net charge density at all, we now have a residual charge given by

$$\rho' = \rho_0 \cosh\zeta - \frac{\rho_0}{\cosh\zeta} = \rho_0 \sinh\zeta \tanh\zeta \tag{4.3}$$

where ρ_0 is the original density of positive and negative charge (which cancels when $\zeta = 0$). The original current may be written in terms of this charge density and the velocity of motion,

$$J_x = -\rho_0 v = -c\rho_0 \tanh\zeta \tag{4.4}$$

where the minus sign is used because the negative charges were in motion. Therefore,

$$\rho' = -\frac{J_x}{c} \sinh\zeta \tag{4.5}$$

The new current density is then given by

$$J'_x = -(\rho_0 \cosh\zeta)v = -c\rho_0 \cosh\zeta \tanh\zeta = J_x \cosh\zeta \tag{4.6}$$

Time dilation was not a factor in this analysis, because the setup of the problem was such that the velocity of the charges was given directly by the velocity of the moving reference frame. If we have components of velocity orthogonal to that of the reference frame, we have to be a bit more careful. Say, for example, that the initial current component was in the y direction. Transformation to the x-moving reference frame would increase the density of moving charge, just as it did for x-directed current. That would tend to increase the J_y current density by a factor $\cosh\zeta$. However, (3.3) tells us that the y component of their velocity is also decreased by the same factor. This can be thought of as a time dilation effect, since events outside the moving reference frame (the translation of charges in an orthogonal direction) appear to progress more slowly. Thus, orthogonal components of the current density are unaffected by the transformation to the new reference frame.

4.1.3 Four-Current

Combining all of these results, we arrive at the following conclusion,

$$c\rho' = c\rho \cosh\zeta - J_x \sinh\zeta \tag{4.7a}$$

$$J'_x = -c\rho \sinh\zeta + J_x \cosh\zeta \tag{4.7b}$$

$$J'_y = J_y \tag{4.7c}$$

$$J'_z = J_z \tag{4.7d}$$

More succinctly,

$$\begin{pmatrix} c\rho' \\ J'_x \\ J'_y \\ J'_z \end{pmatrix} = \begin{pmatrix} \cosh\zeta & -\sinh\zeta & 0 & 0 \\ -\sinh\zeta & \cosh\zeta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c\rho \\ J_x \\ J_y \\ J_z \end{pmatrix} \tag{4.8}$$

which we recognize immediately as a standard Lorentz boost in the x direction. We may thus conclude that the *four-current density*, as the above vector is called, is a bona fide four-vector that transforms according to the same rules as spacetime position and four-velocity. Notably, the four-current density may be written as an equivalent rest charge density, ρ_0, times the four-velocity of those charges,

$$\mathbf{J} = \rho_0 \mathbf{U} \tag{4.9}$$

4.2 RICCI CALCULUS

It is useful at this juncture to begin familiarizing ourselves with some notational conventions that are not often used in classical, nonrelativistic electromagnetics, namely, *Ricci calculus* [1]. We will think of the four-vectors we have described as indexed arrays of numbers, or tensors [2]. For example,

$$X^\mu = \left(X^0, X^1, X^2, X^3\right) = (ct, x, y, z) \tag{4.10a}$$

$$U^\mu = \left(U^0, U^1, U^2, U^3\right) = (c, v_x, v_y, v_z)\cosh\zeta \tag{4.10b}$$

$$K^\mu = \left(K^0, K^1, K^2, K^3\right) = \left(\tfrac{\omega}{c}, k_x, k_y, k_z\right) \tag{4.10c}$$

$$J^\mu = \left(J^0, J^1, J^2, J^3\right) = (c\rho, J_x, J_y, J_z) \tag{4.10d}$$

By convention, a Greek-letter index (such as μ above) shall be understood to range over all four spacetime axes (0–3), while a Roman-letter index would indicate that only the spatial components (1–3) are included.

The position of the index as a superscript is not intended in this case to represent exponentiation. We shall distinguish between tensors having upper and lower indices (also called *contravariant* and *covariant*, respectively) as two different forms of the relevant four-vectors, both transforming between reference frames according to the Lorentz equations (with some sign changes, to be explained shortly). Tensors having one index are called first-rank, while those having two indices are called second-rank, and so on. Examples of second-rank tensors include the Minkowski metric,

$$\eta_{\mu\nu} = \boldsymbol{\eta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{4.11}$$

and the x-directed Lorentz boost operator,

$$L^\mu{}_\nu = \mathbf{L} = \begin{pmatrix} \cosh\zeta & -\sinh\zeta & 0 & 0 \\ -\sinh\zeta & \cosh\zeta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.12}$$

Note that the metric is shown with two lower (covariant) indices, while the boost operator has mixed indices.

4.2.1 Tensor Multiplication

The *Einstein summation rule* says that a tensor product having the same index in both covariant and contravariant positions implies summation over that index. This is called contracting the indices. For example,

$$X'^{\mu} = \Lambda^{\mu}{}_{\nu} X^{\nu} \tag{4.13}$$

could be expanded as

$$X'^{\mu} = \sum_{\nu=0}^{3} \Lambda^{\mu}{}_{\nu} X^{\nu} = \mathbf{\Lambda X} \tag{4.14}$$

or, even more explicitly, as the set of four equations,

$$X'^{0} = \sum_{\nu=0}^{3} \Lambda^{0}{}_{\nu} X^{\nu} = \Lambda^{0}{}_{0} X^{0} + \Lambda^{0}{}_{1} X^{1} + \Lambda^{0}{}_{2} X^{2} + \Lambda^{0}{}_{3} X^{3} \tag{4.15a}$$

$$X'^{1} = \sum_{\nu=0}^{3} \Lambda^{1}{}_{\nu} X^{\nu} = \Lambda^{1}{}_{0} X^{0} + \Lambda^{1}{}_{1} X^{1} + \Lambda^{1}{}_{2} X^{2} + \Lambda^{1}{}_{3} X^{3} \tag{4.15b}$$

$$X'^{2} = \sum_{\nu=0}^{3} \Lambda^{2}{}_{\nu} X^{\nu} = \Lambda^{2}{}_{0} X^{0} + \Lambda^{2}{}_{1} X^{1} + \Lambda^{2}{}_{2} X^{2} + \Lambda^{2}{}_{3} X^{3} \tag{4.15c}$$

$$X'^{3} = \sum_{\nu=0}^{3} \Lambda^{3}{}_{\nu} X^{\nu} = \Lambda^{3}{}_{0} X^{0} + \Lambda^{3}{}_{1} X^{1} + \Lambda^{3}{}_{2} X^{2} + \Lambda^{3}{}_{3} X^{3} \tag{4.15d}$$

This equation should be familiar, representing the transformation of four-position from one reference frame to another.

To convert the index of a tensor from contravariant to covariant, simply multiply by the metric and contract the indices,

$$X_{\mu} = \eta_{\mu\nu} X^{\nu} \tag{4.16}$$

and similarly to convert from covariant to contravariant,

$$X^{\mu} = \eta^{\mu\nu} X_{\nu} \tag{4.17}$$

where $\eta^{\mu\nu} = \eta_{\mu\nu}$. In either case, due to the metric signature we have chosen, raising or lowering the index of a four-vector amounts to nothing more than negating the spacelike components of that four-vector. This immediately leads to the identities

$$\eta^{\mu}{}_{\nu} = \eta^{\mu\sigma} \eta_{\sigma\nu} = \boldsymbol{\eta\eta} = \mathbf{I} \tag{4.18a}$$

$$\eta_{\mu}{}^{\nu} = \eta_{\mu\sigma}\eta^{\sigma\nu} = \boldsymbol{\eta\eta} = \mathbf{I} \tag{4.18b}$$

recalling that $\boldsymbol{\eta}^{-1} = \boldsymbol{\eta}$.

Using the Ricci tensor notation we may rewrite the proper time from (2.49) as

$$(c\tau)^2 = X_{\mu}X^{\mu} \tag{4.19}$$

and the normalization of the four-velocity from (2.60) as

$$U_{\mu}U^{\mu} = c^2 \tag{4.20}$$

or define the Lorentz-invariant phase of a wave from (3.39) as

$$K_{\mu}X^{\mu} = \Phi \tag{4.21}$$

I have committed a slight abuse of notation in some of the examples above by equating the matrix representation of tensors with the tensors themselves, which are, in reality, a different kind of object. Tensors of rank 2 and below can be represented most easily as matrices and column vectors, and the rules for multiplication are similar, with some caveats: The dimensions across which products are summed for matrices are determined by the order of the product elements (across the columns of the first element and down the rows of the second), whereas for tensors, sums are always taken across the indexed dimensions having common labels, regardless of which appears first in the expression. These two notational conventions can be made commensurate so that a direct substitution is allowed if we adopt the following conventions:

- Always write the matrix equivalent of a second-rank tensor so that the first index, whether upper or lower, is a row index, and the second is a column index.
- Reorder the terms of a tensor product so that summed indices always appear in column-then-row order.
- Conventionally, first-rank tensors are represented as column vectors, in which case the lone index is a row index. If a column index is needed for matrix multiplication, the vector should be transposed to produce a row vector.

For example, suppose that we have the following tensor equation,

$$F'^{\mu\nu} = \Lambda^{\mu}{}_{\sigma}\Lambda^{\nu}{}_{\tau}F^{\sigma\tau} \tag{4.22}$$

and that we wish to write this in matrix form. There are two indices of summation, σ and τ, appearing in both covariant and contravariant positions within the expression on the right side. We note that σ appears as the column index of $\Lambda^{\mu}{}_{\sigma}$ and as the row index of $F^{\sigma\tau}$. Since the multiplication of tensors is governed by index labeling, the product terms are commutative (unlike matrices, in general). We may therefore rearrange the terms so that these elements are adjacent to one another,

$$F'^{\mu\nu} = \Lambda^{\mu}{}_{\sigma} F^{\sigma\tau} \Lambda^{\nu}{}_{\tau} \tag{4.23}$$

The next summation index, τ, appears as a column index twice. We can fix that by exchanging the order of the terms in the second operator, $\Lambda^{\nu}{}_{\tau} \to \Lambda_{\tau}{}^{\nu}$. This amounts to the transposition of the equivalent matrix representation. Thus,

$$F'^{\mu\nu} = \Lambda^{\mu}{}_{\sigma} F^{\sigma\tau} \Lambda_{\tau}{}^{\nu} = \mathbf{\Lambda F \Lambda}^{T} \tag{4.24}$$

I will continue to use italic characters with upper and lower indices for tensor notation, but will switch to upright, boldface characters when the matrix representation, and especially matrix product and commutivity rules, are to be applied.

The matrix representation of tensors becomes less effective for rank higher than 2. The back-to-back Lorentz operators in (4.22), for example, could be written as a fourth-rank tensor,

$$\Lambda^{\mu}{}_{\sigma} \Lambda^{\nu}{}_{\tau} = G^{\mu}{}_{\sigma}{}^{\nu}{}_{\tau} \tag{4.25}$$

but to write this as a large matrix would require additional conventions defining how the elements are to be arranged (and the tensor/matrix $F^{\sigma\tau} = \mathbf{F}$ with which it is to be multiplied would have to be expanded in some way to match its oversized dimensions). Tensors in general should be thought of purely as n-dimensional arrays of numbers, for any rank n. Fortunately, such entities are rarely needed in practice.

We now have the tools we need to convert a Lorentz operator (either a rotation, or a boost, or some combination) that transforms contravariant four-vectors, into a form that correctly transforms covariant four-vectors instead,

$$X'^{\mu} = \Lambda^{\mu}{}_{\nu} X^{\nu} \tag{4.26a}$$

$$\therefore \left(\eta^{\mu\sigma} X'_{\sigma}\right) = \Lambda^{\mu}{}_{\nu} \left(\eta^{\nu\tau} X_{\tau}\right) \tag{4.26b}$$

$$X'_{\sigma} = \eta_{\sigma\mu} \Lambda^{\mu}{}_{\nu} \eta^{\nu\tau} X_{\tau} \tag{4.26c}$$

$$\therefore \Lambda'_{\sigma}{}^{\tau} = \eta_{\sigma\mu} \Lambda^{\mu}{}_{\nu} \eta^{\nu\tau} = \boldsymbol{\eta \Lambda \eta} \tag{4.26d}$$

In other words, given the metric signature (+ – – –), we negate the first row and the first column of the original transformation matrix, leaving the principal diagonal element unchanged,

$$\begin{pmatrix} L_{00} & L_{01} & L_{02} & L_{03} \\ L_{10} & L_{11} & L_{12} & L_{13} \\ L_{20} & L_{21} & L_{22} & L_{23} \\ L_{30} & L_{31} & L_{32} & L_{33} \end{pmatrix} \to \begin{pmatrix} L_{00} & -L_{01} & -L_{02} & -L_{03} \\ -L_{10} & L_{11} & L_{12} & L_{13} \\ -L_{20} & L_{21} & L_{22} & L_{23} \\ -L_{30} & L_{31} & L_{32} & L_{33} \end{pmatrix} \tag{4.27}$$

4.2.2 Differential Forms

Differential operators can be written with contravariant and covariant indices as well, such as the covariant *four-gradient*,

$$\partial_\mu = \left(\frac{\partial}{\partial X^0}, \frac{\partial}{\partial X^1}, \frac{\partial}{\partial X^2}, \frac{\partial}{\partial X^3}\right) = \left(\frac{1}{c}\frac{\partial}{\partial t}, \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) \tag{4.28}$$

and the contravariant form [3],

$$\partial^\mu = \eta^{\mu\nu}\partial_\nu = \left(\frac{1}{c}\frac{\partial}{\partial t}, -\frac{\partial}{\partial x}, -\frac{\partial}{\partial y}, -\frac{\partial}{\partial z}\right) \tag{4.29}$$

These are sometimes written more compactly as

$$\partial_\mu = \left(\frac{1}{c}\frac{\partial}{\partial t}, \boldsymbol{\nabla}\right) \tag{4.30a}$$

$$\partial^\mu = \left(\frac{1}{c}\frac{\partial}{\partial t}, -\boldsymbol{\nabla}\right) \tag{4.30b}$$

Note that contracting indices between two four-vectors is a kind of inner product. Thus, contracting the indices of a four-gradient and another four-vector amounts to taking that vector's *four-divergence*. For example, if we take the four-divergence of the four-current density,

$$\partial_\mu J^\mu = \frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{J} \tag{4.31}$$

we discover a compact way to write the continuity equation, (1.15),

$$\partial_\mu J^\mu = 0 \tag{4.32}$$

Interestingly, taking the four-divergence of a four-position (any four-position) simply counts the number of dimensions in spacetime [4],

$$\partial_\mu X^\mu = \frac{1}{c}\frac{\partial ct}{\partial t} + \frac{\partial x}{\partial x} + \frac{\partial y}{\partial y} + \frac{\partial z}{\partial z} = 4 \tag{4.33}$$

Finally, we may note that the d'Alembertian operator can be written quite simply as the inner product of the contravariant and covariant forms of the four-gradient,

$$\partial_\mu \partial^\mu = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2 = \square^2 \tag{4.34}$$

4.2.3 Four-Gradient Transformation

The question remains, is the four-gradient a valid four-vector, in this case a four-vector operator? What I mean is, does the four-gradient operator form transform the same way that an actual four-vector does? To work that out, let us apply the four-gradient to an arbitrary scalar field, say the phase of a wave, Φ,

$$\partial_\mu \Phi = \left(\frac{1}{c}\frac{\partial \Phi}{\partial t}, \frac{\partial \Phi}{\partial x}, \frac{\partial \Phi}{\partial y}, \frac{\partial \Phi}{\partial z}\right) \tag{4.35}$$

As Φ is a Lorentz-invariant scalar, we expect its point-by-point value to be unchanged from Lorentz transformation,

$$\Phi\left(X^\mu\right) = \Phi\left(X'^\mu\right) \tag{4.36}$$

It is helpful if we think of Φ as a function of spacetime coordinates X^μ which in turn are a function of the transformed coordinates X'^ν,

$$\Phi = \Phi\left(X^\mu\left(X'^\nu\right)\right) \tag{4.37}$$

We may then evaluate the derivatives by application of the chain rule [3],

$$\partial'_\nu \Phi = \frac{\partial \Phi}{\partial X'^\nu} = \sum_\mu \frac{\partial \Phi}{\partial X^\mu}\frac{\partial X^\mu}{\partial X'^\nu} = \partial_\mu \Phi \frac{\partial X^\mu}{\partial X'^\nu} \tag{4.38}$$

Since X'^{ν} is just a transformed version of the four-position, X^{μ}, we may recognize the final term above as another way of writing the (covariant) Lorentz transformation tensor,

$$\frac{\partial X^{\mu}}{\partial X'^{\nu}} = \frac{\partial}{\partial X'^{\nu}} \left(X'^{\nu} \Lambda'_{\nu}{}^{\mu} \right) = \Lambda'_{\nu}{}^{\mu} \tag{4.39}$$

Thus,

$$\partial'_{\nu} \Phi = \Lambda'_{\nu}{}^{\mu} \partial_{\mu} \Phi \tag{4.40}$$

or, more simply,

$$\partial'_{\nu} = \Lambda'_{\nu}{}^{\mu} \partial_{\mu} \tag{4.41}$$

which is the exact form for transformation of a covariant four-vector.

Although the choice of invariant scalar field was arbitrary in the above derivation, it so happens that the gradient of the phase returns the four-wavevector of the wave,

$$\partial_{\mu} \Phi = \partial_{\mu} \left(K_{\nu} X^{\nu} \right) = \begin{pmatrix} \frac{1}{c} \frac{\partial}{\partial t} \\ \boldsymbol{\nabla} \end{pmatrix} (\omega t - \mathbf{k} \cdot \mathbf{r}) = \begin{pmatrix} \omega / c \\ -\mathbf{k} \end{pmatrix} = K_{\mu} \tag{4.42}$$

Therefore, (4.40) could just as well have been written as,

$$K'_{\nu} = \Lambda'_{\nu}{}^{\mu} K_{\mu} \tag{4.43}$$

which is true by definition.

4.3 RELATIVISTIC REPRESENTATIONS OF THE ELECTROMAGNETIC FIELD

Having now established, through kinematic arguments, how charges and currents intermix with one another in different inertial reference frames, and having developed a mathematical framework to describe the transformation of these quantities, we would like to rewrite our laws of electromagnetics in a form that is compatible with that framework. In this way, we prepare ourselves to write solutions to electromagnetic problems in a manner that can also be easily transformed from one reference frame to another.

4.3.1 Four-Potential

Recall (1.35), the potential formulation of Maxwell's equations under the Lorenz gauge condition,

$$\Box^2 \varphi = \frac{\rho}{\varepsilon_0} \tag{4.44a}$$

$$\Box^2 \mathbf{A} = \mu_0 \mathbf{J} \tag{4.44b}$$

If we move some constants to the left sides of these equations,

$$\frac{1}{\mu_0}\Box^2 \left(\frac{\varphi}{c}\right) = c\rho \tag{4.45a}$$

$$\frac{1}{\mu_0}\Box^2 \mathbf{A} = \mathbf{J} \tag{4.45b}$$

we see that the right-hand sides, taken together, now represent the relativistic four-current. We may thus take the elements of the left-hand sides as another four-vector, with the d'Alembertian written in four-gradient form,

$$\partial_\mu \partial^\mu A^\nu = \mu_0 J^\nu \tag{4.46}$$

where the tensor A^ν, known as the electromagnetic *four-potential*, is

$$A^\nu = \left(\tfrac{\varphi}{c}, A_x, A_y, A_z\right) \tag{4.47}$$

Transformation of the four-potential should be the same as any other four-vector. This is easily verified by substituting a transformed four-potential into (4.46),

$$\partial_\mu \partial^\mu A'^\nu = \partial_\mu \partial^\mu \left(\Lambda^\nu{}_\sigma A^\sigma\right) = \Lambda^\nu{}_\sigma \left(\partial_\mu \partial^\mu A^\sigma\right) = \Lambda^\nu{}_\sigma \left(\mu_0 J^\sigma\right) \tag{4.48a}$$

$$= \mu_0 \left(\Lambda^\nu{}_\sigma J^\sigma\right) = \mu_0 J'^\nu \tag{4.48b}$$

verifying that Maxwell's equations hold in the new reference frame with the transformed four-current and four-potential.

The Lorenz gauge condition itself may also be written in terms of the four-potential as follows,

$$\partial_\mu A^\mu = \frac{1}{c^2}\frac{\partial \varphi}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{A} = 0 \tag{4.49}$$

Although no physical experiment can distinguish the field potentials constrained by one gauge condition from those of another, the Lorenz gauge is unique among them

Table 4.1
Four-Vectors That Transform by the Lorentz Equations

Name	Symbol	Components				Key Relations
		$\mu = 0$	$\mu = 1$	$\mu = 2$	$\mu = 3$	
Four-position	X^μ	ct	x	y	z	$(c\tau)^2 = X_\mu X^\mu$
Four-velocity*	U^μ	$c\gamma$	$v_x\gamma$	$v_y\gamma$	$v_z\gamma$	$U^\mu = \frac{dX^\mu}{d\tau}$
Four-momentum	P^μ	E/c	p_x	p_y	p_z	$P^\mu = mU^\mu$
Four-force*	F^μ	$\dot{E}\gamma/c$	$\dot{p}_x\gamma$	$\dot{p}_y\gamma$	$\dot{p}_z\gamma$	$F^\mu = \frac{dP^\mu}{d\tau}$
Four-wavevector	K^μ	ω/c	k_x	k_y	k_z	$K_\mu X^\mu = \Phi$
Four-current	J^μ	$c\rho$	J_x	J_y	J_z	$J^\mu = \rho_0 U^\mu$
Four-potential	A^μ	φ/c	A_x	A_y	A_z	$\partial_\nu \partial^\nu A^\mu = \mu_0 J^\mu$

$^*\gamma = \cosh\zeta$.

for being Lorentz invariant [5]. That is, if the Lorenz gauge condition holds true for potentials in one reference frame, it will hold true for the transformed potentials in any other reference frame as well, but the same will not generally be true of, say, the Coulomb gauge condition. (Potentials which originally obeyed the Coulomb gauge, once transformed, will still be valid potentials, they simply will not obey the Coulomb gauge condition in the transformed reference frame.)

A list of all the key four-vectors that we have described thus far can be found in Table 4.1. The electric and magnetic fields themselves cannot be written as four-vectors. That requires another kind of structure entirely, which we discuss in the following section.

4.3.2 The Electromagnetic Field Tensor

If we could write the electric and magnetic fields in terms of the four-potential, which we already know how to transform between reference frames, then we would have a route to determining how to transform the fields as well. We have already proven that the electric and magnetic fields cannot both be written as components of four-vectors in Section 2.1.3, for one would have been an axial vector and the other a polar vector, transforming in different ways.

Let us start with (1.23) for $\mathbf{B}$, repeated here for convenience,

$$\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A} \tag{4.50}$$

or, in expanded form,

$$B_x = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \tag{4.51a}$$

$$B_y = \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \tag{4.51b}$$

$$B_z = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \tag{4.51c}$$

This is a first-order spatial derivative. How do we replicate this in the language of four-vectors and tensors? We might be tempted to try $\partial_\mu A^\mu$, but that has the form of a *dot product* (divergence), not a cross product (curl). Moreover, the Lorenz gauge condition, given in (4.49), guarantees that $\partial_\mu A^\mu = 0$. Let us write $\partial^\mu A^\nu$ instead. Since this expression has no common indices in the covariant and contravariant positions, the summation rule is not activated, and we end up with a second-rank tensor,

$$\partial^\mu A^\nu = \begin{pmatrix} \frac{1}{c}\frac{\partial}{\partial t} \\ -\frac{\partial}{\partial x} \\ -\frac{\partial}{\partial y} \\ -\frac{\partial}{\partial z} \end{pmatrix} \begin{pmatrix} \frac{\varphi}{c} & A_x & A_y & A_z \end{pmatrix} \tag{4.52a}$$

$$= \begin{pmatrix} \frac{1}{c^2}\frac{\partial \varphi}{\partial t} & \frac{1}{c}\frac{\partial A_x}{\partial t} & \frac{1}{c}\frac{\partial A_y}{\partial t} & \frac{1}{c}\frac{\partial A_z}{\partial t} \\ -\frac{1}{c}\frac{\partial \varphi}{\partial x} & -\frac{\partial A_x}{\partial x} & -\frac{\partial A_y}{\partial x} & -\frac{\partial A_z}{\partial x} \\ -\frac{1}{c}\frac{\partial \varphi}{\partial y} & -\frac{\partial A_x}{\partial y} & -\frac{\partial A_y}{\partial y} & -\frac{\partial A_z}{\partial y} \\ -\frac{1}{c}\frac{\partial \varphi}{\partial z} & -\frac{\partial A_x}{\partial z} & -\frac{\partial A_y}{\partial z} & -\frac{\partial A_z}{\partial z} \end{pmatrix} \tag{4.52b}$$

Some of these terms look promising; $-\partial A_x/\partial y$ in row 3, column 2, for example, looks like the second half of the expression for B_z. The first half, $\partial A_y/\partial x$, appears in row 2 column 3, negated. We may complete the curl expressions, then, by subtracting the transpose,

$$\partial^\mu A^\nu - \partial^\nu A^\mu =$$

$$\begin{pmatrix} 0 & \frac{1}{c}\frac{\partial A_x}{\partial t} + \frac{1}{c}\frac{\partial \varphi}{\partial x} & \frac{1}{c}\frac{\partial A_y}{\partial t} + \frac{1}{c}\frac{\partial \varphi}{\partial y} & \frac{1}{c}\frac{\partial A_z}{\partial t} + \frac{1}{c}\frac{\partial \varphi}{\partial z} \\ -\frac{1}{c}\frac{\partial \varphi}{\partial x} - \frac{1}{c}\frac{\partial A_x}{\partial t} & 0 & \frac{\partial A_x}{\partial y} - \frac{\partial A_y}{\partial x} & \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \\ -\frac{1}{c}\frac{\partial \varphi}{\partial y} - \frac{1}{c}\frac{\partial A_y}{\partial t} & \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} & 0 & \frac{\partial A_y}{\partial z} - \frac{\partial A_z}{\partial y} \\ -\frac{1}{c}\frac{\partial \varphi}{\partial z} - \frac{1}{c}\frac{\partial A_z}{\partial t} & \frac{\partial A_z}{\partial x} - \frac{\partial A_x}{\partial z} & \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} & 0 \end{pmatrix} \tag{4.53}$$

We may now recognize many of these terms as being the curl components that identify the magnetic flux density, $\mathbf{B}$,

$$\partial^\mu A^\nu - \partial^\nu A^\mu =$$
$$\begin{pmatrix} 0 & \frac{1}{c}\frac{\partial A_x}{\partial t} + \frac{1}{c}\frac{\partial \varphi}{\partial x} & \frac{1}{c}\frac{\partial A_y}{\partial t} + \frac{1}{c}\frac{\partial \varphi}{\partial y} & \frac{1}{c}\frac{\partial A_z}{\partial t} + \frac{1}{c}\frac{\partial \varphi}{\partial z} \\ -\frac{1}{c}\frac{\partial \varphi}{\partial x} - \frac{1}{c}\frac{\partial A_x}{\partial t} & 0 & -B_z & B_y \\ -\frac{1}{c}\frac{\partial \varphi}{\partial y} - \frac{1}{c}\frac{\partial A_y}{\partial t} & B_z & 0 & -B_x \\ -\frac{1}{c}\frac{\partial \varphi}{\partial z} - \frac{1}{c}\frac{\partial A_z}{\partial t} & -B_y & B_x & 0 \end{pmatrix} \tag{4.54}$$

Moreover, when we consider (1.27) for $\mathbf{E}$,

$$\mathbf{E} = -\boldsymbol{\nabla}\varphi - \frac{\partial \mathbf{A}}{\partial t} \tag{4.55a}$$

$$\therefore E_x = -\frac{\partial \varphi}{\partial x} - \frac{\partial A_x}{\partial t} \tag{4.55b}$$

$$E_y = -\frac{\partial \varphi}{\partial y} - \frac{\partial A_y}{\partial t} \tag{4.55c}$$

$$E_x = -\frac{\partial \varphi}{\partial z} - \frac{\partial A_z}{\partial t} \tag{4.55d}$$

we see that the remaining terms correspond to the electric field,

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix} \tag{4.56}$$

Known as the *electromagnetic field tensor* or the *Faraday tensor*, $F^{\mu\nu}$ is second-rank and antisymmetric, thus having only six independent terms, the exact number of $\mathbf{E}$ and $\mathbf{B}$ components [6, 7]. It can be written more compactly if we define a skew-symmetric matrix expansion for the classical magnetic field vector, $\mathbf{B}$,

$$[\mathbf{B}]_\times = \begin{pmatrix} 0 & -B_z & B_y \\ B_z & 0 & -B_x \\ -B_y & B_x & 0 \end{pmatrix} \tag{4.57}$$

Note that the format of $[\mathbf{B}]_\times$ makes it easy to write cross products and curls; for any three-vectors, $\mathbf{a}$ and $\mathbf{b}$, the cross product is simply $[\mathbf{a}]_\times \mathbf{b}$, as illustrated in Section B.1.2. Thus, the electromagnetic field tensor in block matrix form is

$$F^{\mu\nu} = \begin{pmatrix} 0 & -\frac{1}{c}\mathbf{E}^T \\ \frac{1}{c}\mathbf{E} & [\mathbf{B}]_\times \end{pmatrix} = \frac{1}{c}\begin{pmatrix} 0 & -\mathbf{E}^T \\ \mathbf{E} & \eta_0\,[\mathbf{H}]_\times \end{pmatrix} \tag{4.58}$$

where $\eta_0 = \sqrt{\mu_0/\varepsilon_0} \approx 377\Omega$ is the *wave impedance* of free space.

The indices as shown are contravariant. Sometimes it will be convenient to work with covariant indices, which can be found using (4.27),

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{pmatrix} \tag{4.59a}$$

$$= \begin{pmatrix} 0 & \frac{1}{c}\mathbf{E}^T \\ -\frac{1}{c}\mathbf{E} & [\mathbf{B}]_\times \end{pmatrix} = \frac{1}{c}\begin{pmatrix} 0 & \mathbf{E}^T \\ -\mathbf{E} & \eta_0\,[\mathbf{H}]_\times \end{pmatrix} \tag{4.59b}$$

The classical electric and magnetic field three-vectors may be recovered from the tensor form as follows [8],

$$\mathbf{E} = E_i = cF_{0i} \tag{4.60a}$$

$$\mathbf{H} = H_i = -\frac{1}{2\mu_0}\epsilon_{ijk}F^{jk} \tag{4.60b}$$

where we may recall that Roman-letter indices (such as i, j, and k which appear in these equations) span only the spatial components, 1–3, and ϵ_{ijk} is the third-rank *Levi-Civita tensor*. The Levi-Civita tensor (or sometimes the *Levi-Civita symbol*) has value given by

$$\epsilon_{ijk\ldots} = \begin{cases} +1 & \text{if } (i,j,k,\ldots) \text{ is an even permutation of } (1,2,3\ldots) \\ -1 & \text{if } (i,j,k,\ldots) \text{ is an odd permutation of } (1,2,3\ldots) \\ 0 & \text{otherwise} \end{cases} \tag{4.61}$$

Thus, for example,

$$\mathbf{H} = -\frac{1}{2\mu_0}\begin{pmatrix} \epsilon_{123}F^{23} + \epsilon_{132}F^{32} \\ \epsilon_{231}F^{31} + \epsilon_{213}F^{13} \\ \epsilon_{312}F^{12} + \epsilon_{321}F^{21} \end{pmatrix} = -\frac{1}{2\mu_0}\begin{pmatrix} F^{23} - F^{32} \\ F^{31} - F^{13} \\ F^{12} - F^{21} \end{pmatrix} \tag{4.62a}$$

$$= \frac{1}{2\mu_0} \begin{pmatrix} 2B_x \\ 2B_y \\ 2B_z \end{pmatrix} = \begin{pmatrix} H_x \\ H_y \\ H_z \end{pmatrix} \tag{4.62b}$$

The field tensor also has a dual form given by

$$G^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\sigma\tau} F_{\sigma\tau} = \begin{pmatrix} 0 & -B_x & -B_y & -B_z \\ B_x & 0 & E_z/c & -E_y/c \\ B_y & -E_z/c & 0 & E_x/c \\ B_z & E_y/c & -E_x/c & 0 \end{pmatrix} \tag{4.63a}$$

$$= \begin{pmatrix} 0 & -\mathbf{B}^T \\ \mathbf{B} & -\frac{1}{c}\left[\mathbf{E}\right]_\times \end{pmatrix} = \frac{1}{c}\begin{pmatrix} 0 & -\eta_0 \mathbf{H}^T \\ \eta_0 \mathbf{H} & -\left[\mathbf{E}\right]_\times \end{pmatrix} \tag{4.63b}$$

where $\epsilon^{\mu\nu\sigma\tau}$ is the fourth-rank Levi-Civita symbol. This is known as the *Hodge dual* of the field tensor $F^{\mu\nu}$, and is written compactly using the *Hodge star operator*,[1]

$$G^{\mu\nu} = \star F^{\mu\nu} \tag{4.64}$$

Some important equations can be expressed more simply using the dual field tensor. For quick reference, the common electromagnetic field tensor forms are listed in Table 4.2.

4.3.3 Field Tensor Invariants

It is worth noting that there are several Lorentz-invariant scalar quantities that can be associated with the electromagnetic field tensor [8]. The first is the inner product of the covariant and contravariant forms,

$$F_{\mu\nu}F^{\mu\nu} = 2\left(B^2 - \tfrac{1}{c^2}E^2\right) \tag{4.65}$$

We may also take the inner product of the field tensor with its dual,

$$G_{\mu\nu}F^{\mu\nu} = -\tfrac{4}{c}\mathbf{B}\cdot\mathbf{E} \tag{4.66}$$

This is a pseudoscalar quantity (see Section 2.1.3) which is invariant under Lorentz transformation. Finally, the determinant of the field tensor itself,

$$\det F^{\mu\nu} = \tfrac{1}{c^2}\left(\mathbf{B}\cdot\mathbf{E}\right)^2 = \tfrac{1}{16}\left(G_{\mu\nu}F^{\mu\nu}\right)^2 \tag{4.67}$$

1 The Hodge star operator actually has a much broader definition than is implied in (4.63a), but this will suffice for its application to second-rank tensors in this book [9].

Table 4.2
Electromagnetic Field Tensors (+ − − −)

Description	Symbol	Block Matrix	Elements
Contravariant	$F^{\mu\nu} =$	$\begin{pmatrix} 0 & -\frac{1}{c}\mathbf{E}^T \\ \frac{1}{c}\mathbf{E} & [\mathbf{B}]_\times \end{pmatrix} =$	$\begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix}$
Covariant	$F_{\mu\nu} =$	$\begin{pmatrix} 0 & \frac{1}{c}\mathbf{E}^T \\ -\frac{1}{c}\mathbf{E} & [\mathbf{B}]_\times \end{pmatrix} =$	$\begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{pmatrix}$
Dual, contravariant	$G^{\mu\nu} =$	$\begin{pmatrix} 0 & -\mathbf{B}^T \\ \mathbf{B} & -\frac{1}{c}[\mathbf{E}]_\times \end{pmatrix} =$	$\begin{pmatrix} 0 & -B_x & -B_y & -B_z \\ B_x & 0 & E_z/c & -E_y/c \\ B_y & -E_z/c & 0 & E_x/c \\ B_z & E_y/c & -E_x/c & 0 \end{pmatrix}$
Dual, covariant	$G_{\mu\nu} =$	$\begin{pmatrix} 0 & \mathbf{B}^T \\ -\mathbf{B} & -\frac{1}{c}[\mathbf{E}]_\times \end{pmatrix} =$	$\begin{pmatrix} 0 & B_x & B_y & B_z \\ -B_x & 0 & E_z/c & -E_y/c \\ -B_y & -E_z/c & 0 & E_x/c \\ -B_z & E_y/c & -E_x/c & 0 \end{pmatrix}$

is invariant (as it should be, since the determinant of any Lorentz transformation operator is always 1).

4.3.4 Transformation of the Field Tensor

Having now expressed the electromagnetic field tensor in terms of a four-vector quantity (the four-potential), we can derive an expression for its transformation properties,

$$F'^{\mu\nu} = \partial'^{\mu} A'^{\nu} - \partial'^{\nu} A'^{\mu} = \left(\Lambda^{\mu}{}_{\sigma}\partial^{\sigma}\right)\left(\Lambda^{\nu}{}_{\tau}A^{\tau}\right) - \left(\Lambda^{\nu}{}_{\tau}\partial^{\tau}\right)\left(\Lambda^{\mu}{}_{\sigma}A^{\sigma}\right) \tag{4.68a}$$

$$= \Lambda^{\mu}{}_{\sigma}\Lambda^{\nu}{}_{\tau}\left(\partial^{\sigma}A^{\tau} - \partial^{\tau}A^{\sigma}\right) = \Lambda^{\mu}{}_{\sigma}\Lambda^{\nu}{}_{\tau}F^{\sigma\tau} \tag{4.68b}$$

Figure 4.3 Transformation of the electromagnetic field tensor. (a) An electron in a y-directed wire moving in the x direction through a z-directed magnetic field, $\mathbf{B}$, experiences a force in the positive-y direction. (b) The same force in a reference frame that is comoving with the wire must be due to an electric field, $\mathbf{E}$, instead, since the electron starts apparently at rest.

This is the example we worked out in Section 4.2.1, with the solution as follows,

$$F'^{\mu\nu} = \Lambda^{\mu}{}_{\sigma}\Lambda^{\nu}{}_{\tau}F^{\sigma\tau} = \Lambda^{\mu}{}_{\sigma}F^{\sigma\tau}\Lambda_{\tau}{}^{\nu} = \mathbf{\Lambda F \Lambda}^{T} \tag{4.69}$$

Example: Let us apply this to an example posed by Einstein himself (though my presentation will differ slightly from his in the details to make it more familiar to the modern engineering practitioner) [3, 10]. Imagine a wire suspended in the y direction, and being moved in the x-direction in a laboratory reference frame through a z-directed magnetic field. This is illustrated in Figure 4.3(a). Electrons in the wire will therefore experience a force due to the Lorentz force law given by the cross product of their velocity (the velocity of the wire) and the ambient magnetic field. Since the electrons' charge is negative, that force, and the subsequent acceleration, is in the positive-y direction.

However, if we transform to a comoving reference frame with the wire, the electrons within it are initially at rest. How then can they experience a force and acceleration in the positive-y direction? The only force a stationary charge can experience is from an electric field. We are forced to conclude that there is an electric field in the comoving reference frame, which was not present in the laboratory reference frame.

Let us see if our tensor field formalism predicts this. The setup of our problem in the laboratory reference frame can be summarized with the following equations,

$$\mathbf{v} = v_x \mathbf{x} \tag{4.70a}$$

$$\mathbf{E} = 0 \tag{4.70b}$$

$$\mathbf{B} = B_z \mathbf{z} \tag{4.70c}$$

The field tensor, then, is

$$F^{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -B_z & 0 \\ 0 & B_z & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{4.71}$$

and the force on an electron in the wire is given by (1.10),

$$\mathbf{F}_m = q\mathbf{v} \times \mathbf{B} = e_c v_x B_z \mathbf{y} \tag{4.72}$$

where we have assigned the electronic charge $q = -e_c$. Transformation into the comoving reference frame proceeds as given in (4.69). First, we write the boost operator for motion in the x direction,

$$L^{\mu}{}_{\sigma} = L^{\nu}{}_{\tau} = \mathbf{L} = \begin{pmatrix} \cosh\zeta & -\sinh\zeta & 0 & 0 \\ -\sinh\zeta & \cosh\zeta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.73}$$

where $c\tanh\zeta = v_x$. The transformed field tensor is then found,

$$F'^{\mu\nu} = \begin{pmatrix} 0 & -E'_x/c & -E'_y/c & -E'_z/c \\ E'_x/c & 0 & -B'_z & B'_y \\ E'_y/c & B'_z & 0 & -B'_x \\ E'_z/c & -B'_y & B'_x & 0 \end{pmatrix} = \Lambda^{\mu}{}_{\sigma} F^{\sigma\tau} \Lambda_{\tau}{}^{\nu} \tag{4.74a}$$

$$= \begin{pmatrix} \cosh\zeta & -\sinh\zeta & 0 & 0 \\ -\sinh\zeta & \cosh\zeta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -B_z & 0 \\ 0 & B_z & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \cosh\zeta & -\sinh\zeta & 0 & 0 \\ -\sinh\zeta & \cosh\zeta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.74b}$$

$$= \begin{pmatrix} 0 & 0 & B_z\sinh\zeta & 0 \\ 0 & 0 & -B_z\cosh\zeta & 0 \\ 0 & B_z & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \cosh\zeta & -\sinh\zeta & 0 & 0 \\ -\sinh\zeta & \cosh\zeta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.74c}$$

$$= \begin{pmatrix} 0 & 0 & B_z\sinh\zeta & 0 \\ 0 & 0 & -B_z\cosh\zeta & 0 \\ -B_z\sinh\zeta & B_z\cosh\zeta & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{4.74d}$$

So, the classical fields in the comoving reference frame are

$$\mathbf{B}' = (B_z \cosh\zeta)\,\mathbf{z} \tag{4.75a}$$

$$\mathbf{E}' = -\,(B_z c \sinh\zeta)\,\mathbf{y} \tag{4.75b}$$

The transformed magnetic field has no immediate impact on the unmoving electrons. However, an electric field has appeared, and it exerts a force given by

$$\mathbf{F}'_e = q\mathbf{E}' = e_c\,(B_z c \sinh\zeta)\,\mathbf{y} = (e_c v_x B_z \mathbf{y})\cosh\zeta = \mathbf{F}_m \cosh\zeta \tag{4.76}$$

It is interesting to note that the electric force here does not exactly match (4.72), the original magnetic force, though they do agree tangentially for small velocity, $v_x \ll c$. The reason for this is time dilation. In the laboratory frame, time for the electron is dilated, so we perceive it as having less acceleration, and hence less force, by a factor of $\cosh\zeta$ compared to the comoving frame. To put it another way, in relativistic terms, force acts on the four-velocity, which is a derivative with respect to proper time, not the laboratory time.

4.3.5 The Joules-Bernoulli Equations

Sometimes it is useful to decompose the field tensor into components which are either longitudinal or transverse to the direction of a boost. For a z-directed boost, the transverse components are

$$F_t^{\mu\nu} = \begin{pmatrix} 0 & -\frac{1}{c}\mathbf{E}_t^T \\ \frac{1}{c}\mathbf{E}_t & [\mathbf{B}_t]_\times \end{pmatrix} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & 0 \\ E_x/c & 0 & 0 & B_y \\ E_y/c & 0 & 0 & -B_x \\ 0 & -B_y & B_x & 0 \end{pmatrix} \tag{4.77}$$

while the longitudinal components are

$$F_z^{\mu\nu} = \begin{pmatrix} 0 & -\frac{1}{c}\mathbf{E}_z^T \\ \frac{1}{c}\mathbf{E}_z & [\mathbf{B}_z]_\times \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & -E_z/c \\ 0 & 0 & -B_z & 0 \\ 0 & B_z & 0 & 0 \\ E_z/c & 0 & 0 & 0 \end{pmatrix} \tag{4.78}$$

Let us further use the block form of the Lorentz boost operator from Table 2.2,

$$L^\mu{}_\nu = \begin{pmatrix} \gamma & -\boldsymbol{\beta}^T\gamma \\ -\boldsymbol{\beta}\gamma & \mathbf{L}_s \end{pmatrix} \tag{4.79}$$

where

$$\mathbf{L}_s = \mathbf{L}_s^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \gamma \end{pmatrix} \tag{4.80a}$$

$$\boldsymbol{\beta} = \frac{\mathbf{v}}{c} = \frac{v}{c}\mathbf{z} = (\tanh\zeta)\mathbf{z} \tag{4.80b}$$

$$\gamma = \cosh\zeta \tag{4.80c}$$

Thus, the transverse components transform as

$$(F_t^{\mu\nu})' = L^{\mu}{}_{\sigma} F_t^{\sigma\tau} L_{\tau}{}^{\nu} \tag{4.81a}$$

$$= \begin{pmatrix} \gamma & -\boldsymbol{\beta}^T\gamma \\ -\boldsymbol{\beta}\gamma & \mathbf{L}_s \end{pmatrix} \begin{pmatrix} 0 & -\frac{1}{c}\mathbf{E}_t^T \\ \frac{1}{c}\mathbf{E}_t & [\mathbf{B}_t]_\times \end{pmatrix} \begin{pmatrix} \gamma & -\boldsymbol{\beta}^T\gamma \\ -\boldsymbol{\beta}\gamma & \mathbf{L}_s \end{pmatrix} \tag{4.81b}$$

$$= \begin{pmatrix} \gamma & -\boldsymbol{\beta}^T\gamma \\ -\boldsymbol{\beta}\gamma & \mathbf{L}_s \end{pmatrix} \begin{pmatrix} 0 & -\frac{1}{c}\mathbf{E}_t^T \\ -\frac{\gamma}{c}\mathbf{B}\times\mathbf{v} + \frac{\gamma}{c}\mathbf{E}_t & [\mathbf{B}_t]_\times \mathbf{L}_s - \frac{\gamma}{c^2}\mathbf{E}_t\otimes\mathbf{v} \end{pmatrix} \tag{4.81c}$$

$$= \begin{pmatrix} 0 & \left(\frac{\gamma}{c}\mathbf{B}\times\mathbf{v} - \frac{\gamma}{c}\mathbf{E}_t\right)^T \\ -\frac{\gamma}{c}\mathbf{B}\times\mathbf{v} + \frac{\gamma}{c}\mathbf{E}_t & \gamma\,[\mathbf{B}_t]_\times - \frac{\gamma}{c^2}\,[\mathbf{v}\times\mathbf{E}]_\times \end{pmatrix} \tag{4.81d}$$

where we have used the identity (B.12) for the commutator product of $\mathbf{E}$ and $\mathbf{v}$. The longitudinal components transform as

$$(F_z^{\mu\nu})' = L^{\mu}{}_{\sigma} F_z^{\sigma\tau} L_{\tau}{}^{\nu} \tag{4.82a}$$

$$= \begin{pmatrix} \gamma & -\boldsymbol{\beta}^T\gamma \\ -\boldsymbol{\beta}\gamma & \mathbf{L}_s \end{pmatrix} \begin{pmatrix} 0 & -\frac{1}{c}\mathbf{E}_z^T \\ \frac{1}{c}\mathbf{E}_z & [\mathbf{B}_z]_\times \end{pmatrix} \begin{pmatrix} \gamma & -\boldsymbol{\beta}^T\gamma \\ -\boldsymbol{\beta}\gamma & \mathbf{L}_s \end{pmatrix} \tag{4.82b}$$

$$= \begin{pmatrix} \gamma & -\boldsymbol{\beta}^T\gamma \\ -\boldsymbol{\beta}\gamma & \mathbf{L}_s \end{pmatrix} \begin{pmatrix} \frac{\gamma}{c^2}\mathbf{E}\cdot\mathbf{v} & -\frac{\gamma}{c}\mathbf{E}_z^T \\ \frac{\gamma}{c}\mathbf{E}_z & [\mathbf{B}_z]_\times - \frac{\gamma}{c^2}\mathbf{E}_z\otimes\mathbf{v} \end{pmatrix} \tag{4.82c}$$

$$= \begin{pmatrix} 0 & -\frac{1}{c}\mathbf{E}_z^T \\ \frac{1}{c}\mathbf{E}_z & [\mathbf{B}_z]_\times \end{pmatrix} \tag{4.82d}$$

Combining these results, we have

$$F'^{\mu\nu} = (F_t^{\mu\nu})' + (F_z^{\mu\nu})' = \begin{pmatrix} 0 & -\frac{1}{c}\mathbf{E}'^T \\ \frac{1}{c}\mathbf{E}' & [\mathbf{B}']_\times \end{pmatrix} \tag{4.83a}$$

$$= \begin{pmatrix} 0 & \left(\frac{\gamma}{c}\mathbf{B}\times\mathbf{v} - \frac{\gamma}{c}\mathbf{E}_t - \frac{1}{c}\mathbf{E}_z\right)^T \\ -\frac{\gamma}{c}\mathbf{B}\times\mathbf{v} + \frac{\gamma}{c}\mathbf{E}_t + \frac{1}{c}\mathbf{E}_z & \left[\gamma\mathbf{B}_t + \mathbf{B}_z\right]_\times - \frac{\gamma}{c^2}\left[\mathbf{v}\times\mathbf{E}\right]_\times \end{pmatrix} \tag{4.83b}$$

Therefore, the transformed fields are given by

$$\mathbf{B}' = \gamma\left(\mathbf{B}_t + c^{-2}\mathbf{E}\times\mathbf{v}\right) + \mathbf{B}_z \tag{4.84a}$$

$$\mathbf{E}' = \gamma\left(\mathbf{E}_t + \mathbf{v}\times\mathbf{B}\right) + \mathbf{E}_z \tag{4.84b}$$

More arbitrarily, for a boost velocity in any direction, $\mathbf{v} = v\mathbf{n}$, we may substitute

$$\mathbf{E}_z \to \mathbf{E}_{\|} = (\mathbf{E}\cdot\mathbf{n})\mathbf{n} \tag{4.85a}$$

$$\mathbf{B}_z \to \mathbf{B}_{\|} = (\mathbf{B}\cdot\mathbf{n})\mathbf{n} \tag{4.85b}$$

$$\mathbf{E}_t \to \mathbf{E}_{\perp} = \mathbf{E} - (\mathbf{E}\cdot\mathbf{n})\mathbf{n} \tag{4.85c}$$

$$\mathbf{B}_t \to \mathbf{B}_{\perp} = \mathbf{B} - (\mathbf{B}\cdot\mathbf{n})\mathbf{n} \tag{4.85d}$$

giving rise to the general expressions

$$\mathbf{B}' = \gamma\left(\mathbf{B} + c^{-2}\mathbf{E}\times\mathbf{v}\right) - (\gamma-1)(\mathbf{B}\cdot\mathbf{n})\mathbf{n} \tag{4.86a}$$

$$\mathbf{E}' = \gamma\left(\mathbf{E} + \mathbf{v}\times\mathbf{B}\right) - (\gamma-1)(\mathbf{E}\cdot\mathbf{n})\mathbf{n} \tag{4.86b}$$

or

$$\mathbf{E}'_{\|} = \mathbf{E}_{\|} \tag{4.87a}$$

$$\mathbf{B}'_{\|} = \mathbf{B}_{\|} \tag{4.87b}$$

$$\mathbf{E}'_{\perp} = \gamma\left(\mathbf{E}_{\perp} + \mathbf{v}\times\mathbf{B}\right) \tag{4.87c}$$

$$\mathbf{B}'_{\perp} = \gamma\left(\mathbf{B}_{\perp} + c^{-2}\mathbf{E}\times\mathbf{v}\right) \tag{4.87d}$$

However they are written, these relationships are called the *Joules-Bernoulli equations*. The transformation of electromagnetic fields is thus summarized in Table 4.3. The last form (4.87) makes it clear that, in contrast to four-vectors, the field tensor components that are directed along the axis of motion are unaffected by the boost, while the transverse components instead are modified.

Example: We may illustrate the application of these rules by considering the simple example of a lone point charge, shown in Figure 4.4(a). The electrostatic field in the case where the particle of charge is stationary is found almost trivially by Coulomb's

Table 4.3
Reference Frame Transformation of Electromagnetic Fields

Description	Expression
General tensor form	$F'^{\mu\nu} = \Lambda^{\mu}{}_{\sigma} F^{\sigma\tau} \Lambda_{\tau}{}^{\nu}$
Joules-Bernoulli equations boost velocity $\mathbf{v} = v\mathbf{n}$	$\mathbf{B}' = \gamma\left(\mathbf{B} + c^{-2}\mathbf{E} \times \mathbf{v}\right) - (\gamma - 1)(\mathbf{B} \cdot \mathbf{n})\mathbf{n}$ $\mathbf{E}' = \gamma\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right) - (\gamma - 1)(\mathbf{E} \cdot \mathbf{n})\mathbf{n}$
Joules-Bernoulli equations parallel-perpendicular decomposition	$\mathbf{E}'_{\|\|} = \mathbf{E}_{\|\|}$ $\mathbf{B}'_{\|\|} = \mathbf{B}_{\|\|}$ $\mathbf{E}'_{\perp} = \gamma\left(\mathbf{E}_{\perp} + \mathbf{v} \times \mathbf{B}\right)$ $\mathbf{B}'_{\perp} = \gamma\left(\mathbf{B}_{\perp} + c^{-2}\mathbf{E} \times \mathbf{v}\right)$

Figure 4.4 Emergence of the magnetic field as a dynamic correction to the fundamental Coulomb force. (a) Electrostatic field of a stationary point charge. (b) Transformation to a reference frame in which the charge is seen to be moving reveals an emergent magnetic field.

law,

$$\mathbf{E} = \frac{q}{4\pi\varepsilon_0 r^2}\mathbf{r} \tag{4.88}$$

where $\mathbf{r}$ is the unit vector in the radial direction. The transverse and longitudinal components may be written separately in terms of the declination angle, θ, from the z axis,

$$\mathbf{E}_{\perp} = E\sin\theta\boldsymbol{\rho} = \frac{q\sin\theta}{4\pi\varepsilon_0 r^2}\boldsymbol{\rho} \tag{4.89a}$$

$$\mathbf{E}_{||} = E\cos\theta\mathbf{z} = \frac{q\cos\theta}{4\pi\varepsilon_0 r^2}\mathbf{z} \tag{4.89b}$$

There is no magnetic field in the electrostatic case. However, if we transform to a reference frame moving at speed v in the negative-z direction, so that the charged particle now appears to be moving in the positive-z direction, as in Figure 4.4(b), we find from the Joules-Bernoulli equations,

$$\mathbf{E}'_\perp = \gamma \mathbf{E}_\perp = \frac{q \sin\theta}{4\pi\varepsilon_0 r^2}(\cosh\zeta)\boldsymbol{\rho} \tag{4.90a}$$

$$\mathbf{B}'_\perp = \frac{\gamma}{c^2}\mathbf{E}\times\mathbf{v} = -\frac{1}{c}\frac{q}{4\pi\varepsilon_0 r^2}(\sinh\zeta)\boldsymbol{\phi} \tag{4.90b}$$

We therefore see that a magnetic field arises directly as a consequence of the fundamental electrostatic force and the principle of *Lorentz covariance* — the requirement that electromagnetic laws have the same form in all inertial reference frames under Lorentz transformations.[2] This is why we say that magnetic fields are a required relativistic correction to electric fields.

4.3.6 The Stress-Energy Tensor

An important quantity in relativistic field analysis is the electromagnetic stress-energy tensor [11, 12],

$$T^{\mu\nu} = -\frac{1}{\mu_0}\left(F^{\mu\sigma}F^{\nu}{}_{\sigma} - \frac{1}{4}\eta^{\mu\nu}F_{\sigma\tau}F^{\sigma\tau}\right) \tag{4.91a}$$

$$= -\frac{1}{\mu_0}\left(\eta_{\sigma\tau}F^{\mu\sigma}F^{\nu\tau} - \frac{1}{4}\eta^{\mu\nu}F_{\sigma\tau}F^{\sigma\tau}\right) \tag{4.91b}$$

$$= -\frac{1}{2\mu_0}\left(F^{\mu}{}_{\sigma}F^{\nu\sigma} + G^{\mu}{}_{\sigma}G^{\nu\sigma}\right) \tag{4.91c}$$

$$= \begin{pmatrix} u & S_x/c & S_y/c & S_z/c \\ S_x/c & -\sigma_{xx} & -\sigma_{xy} & -\sigma_{xz} \\ S_y/c & -\sigma_{yx} & -\sigma_{yy} & -\sigma_{yz} \\ S_z/c & -\sigma_{zx} & -\sigma_{zy} & -\sigma_{zz} \end{pmatrix} = \begin{pmatrix} u & \frac{1}{c}\mathbf{S}^T \\ \frac{1}{c}\mathbf{S} & -\boldsymbol{\sigma} \end{pmatrix} \tag{4.91d}$$

where the spatial quadrant is populated with $\boldsymbol{\sigma}$, the Maxwell stress tensor, $\mathbf{S} = S_x\mathbf{x} + S_y\mathbf{y} + S_z\mathbf{z}$ is the Poynting vector, and u is the total electromagnetic energy density, all described in Section 1.4. The stress-energy tensor has units of energy per volume, or equivalently, pressure. It describes the electromagnetic contribution to total energy and momentum which are the sources of gravitational fields in Einstein's general relativistic field theory, just as mass is the source of gravitation in Newton's inverse-square law of gravity [13].

2 This should not be confused with the covariance property of a four-vector having lower indices. Both covariant and contravariant four-vectors may be Lorentz-covariant quantities appearing in Lorentz-covariant equations. The similarity of naming conventions is an unfortunate coincidence of history.

4.4 MAXWELL'S EQUATIONS IN TENSOR FORM

We will now recast the known laws of electromagnetics into forms that explicitly operate on the four-vectors and tensors we have identified. These equations shall be called Lorentz covariant in that they operate on quantities that transform according to the Lorentz equations, and thus they take on the same form in all inertial reference frames.

Equation (4.56) showed us how to derive the electromagnetic field tensor from the four-potential, but sometimes it is preferable to work directly with the sources themselves, which in this context are represented by the four-current. Such solutions will be our new expression of Maxwell's equations.

4.4.1 The Inhomogeneous Equation

We may eliminate the four-potential by differentiating the field tensor with ∂_μ,

$$\partial_\mu F^{\mu\nu} = \partial_\mu \left(\partial^\mu A^\nu - \partial^\nu A^\mu\right) = \partial_\mu \partial^\mu A^\nu - \partial^\nu \left(\partial_\mu A^\mu\right) \tag{4.92}$$

We may substitute (4.46) for the first term, and note that the part of the second term in parentheses is merely the Lorenz gauge condition (4.49), which we have set to zero. Therefore,

$$\partial_\mu F^{\mu\nu} = \mu_0 J^\nu \tag{4.93}$$

(Alternatively, we could differentiate with respect to the other index, in which case we would have $\partial_\nu F^{\mu\nu} = -\mu_0 J^\mu$.) This simple expression embodies both Gauss's law and Ampère's law, and may be called the *Gauss-Ampère law*. It is also sometimes called the inhomogeneous Maxwell's equation since it contains an undifferentiated source term, J^μ (combining ρ and $\mathbf{J}$ from the classical formulation).

4.4.2 The Homogeneous Equation

However, Faraday's law and the law of magnetic nondivergence (the homogeneous equations) were guaranteed implicitly in the potential formulation by the way the scalar and vector potentials were defined. We need another condition to guarantee those relationships, now that the potentials themselves are absent. That condition is called the Bianchi identity [3],

$$\partial_\mu F_{\nu\sigma} + \partial_\sigma F_{\mu\nu} + \partial_\nu F_{\sigma\mu} = 0 \tag{4.94}$$

To verify that it works, we may substitute for the tensor field from (4.59a),

$$\begin{aligned}&\partial_\mu\left(\partial_\nu A_\sigma-\partial_\sigma A_\nu\right)+\partial_\sigma\left(\partial_\mu A_\nu-\partial_\nu A_\mu\right)+\partial_\nu\left(\partial_\sigma A_\mu-\partial_\mu A_\sigma\right)\\&\quad=\partial_\mu\partial_\nu A_\sigma-\partial_\sigma\partial_\mu A_\nu+\partial_\sigma\partial_\mu A_\nu-\partial_\nu\partial_\sigma A_\mu+\partial_\nu\partial_\sigma A_\mu-\partial_\mu\partial_\nu A_\sigma \quad (4.95a)\\&\quad=0 \quad (4.95b)\end{aligned}$$

For those who obsess over compactness, there is a very short way to write this,

$$\partial_{[\mu}F_{\nu\sigma]}=0 \tag{4.96}$$

where the square brackets denote the *antisymmetrization* of the tensor with respect to the enclosed indices. This means that we average over all signed permutations of the indices (the sign is positive if the number of adjacent transpositions in the permutation is even, and negative if it is odd). Thus,

$$\partial_{[\mu}F_{\nu\sigma]}=\tfrac{1}{6}\left(\partial_\mu F_{\nu\sigma}+\partial_\sigma F_{\mu\nu}+\partial_\nu F_{\sigma\mu}-\partial_\mu F_{\sigma\nu}-\partial_\sigma F_{\nu\mu}-\partial_\nu F_{\mu\sigma}\right) \tag{4.97a}$$

$$=\tfrac{1}{3}\left(\partial_\mu F_{\nu\sigma}+\partial_\sigma F_{\mu\nu}+\partial_\nu F_{\sigma\mu}\right)=0 \tag{4.97b}$$

where we have used the fact that the field tensor is antisymmetric ($F_{\mu\nu}=-F_{\nu\mu}$). This notation is also a useful way to write the field tensor in terms of the four-potential,

$$F^{\mu\nu}=\partial^\mu A^\nu-\partial^\nu A^\mu=2\partial^{[\mu}A^{\nu]} \tag{4.98}$$

Alternatively, the homogeneous Maxwell's equations could be written using the dual field tensor,

$$\partial_\nu G^{\mu\nu}=0 \tag{4.99}$$

because[3]

$$\partial_\nu G^{\mu\nu}=\begin{pmatrix}0 & -\mathbf{B}^T\\ \mathbf{B} & -\frac{1}{c}\left[\mathbf{E}\right]_\times\end{pmatrix}\begin{pmatrix}\frac{1}{c}\frac{\partial}{\partial t}\\ \boldsymbol{\nabla}\end{pmatrix}=\begin{pmatrix}-\boldsymbol{\nabla}\cdot\mathbf{B}\\ \frac{1}{c}\left(\boldsymbol{\nabla}\times\mathbf{E}+\frac{\partial\mathbf{B}}{\partial t}\right)\end{pmatrix}=\mathbf{0} \tag{4.100}$$

revealing Faraday's law and the magnetic nondivergence in the spatial and temporal components of the result, respectively.

3 Matrix multiplication rules require that we put the column vector ∂_ν last, but it should be understood from context that the differential operators within it still operate upon the elements of the preceding field tensor.

Whichever form we use, it is noteworthy that we have reduced Maxwell's equations now from four to two. This is natural since we have combined the electric and magnetic fields into a single quantity, just as the space and time coordinates of events were combined into four-positions, current and charge into the four-current, and wavenumber and frequency into the four-wavevector.

4.5 LORENTZ FORCE LAW IN TENSOR FORM

As long as we have expressed the electric and magnetic fields as a tensor, it would be useful to express (1.10), the Lorentz force law, using the same quantity. Note that the product of the field tensor with the four-velocity,

$$F^{\mu\nu}U_\nu = \begin{pmatrix} 0 & -\frac{1}{c}\mathbf{E}^T \\ \frac{1}{c}\mathbf{E} & [\mathbf{B}]_\times \end{pmatrix} \begin{pmatrix} c \\ -\mathbf{v} \end{pmatrix} \cosh\zeta = \begin{pmatrix} \mathbf{v}\cdot\mathbf{E}/c \\ \mathbf{E}+\mathbf{v}\times\mathbf{B} \end{pmatrix} \cosh\zeta \tag{4.101}$$

has the right form in the spatial components. We may thus write the force law as

$$\frac{dP^\mu}{d\tau} = m\frac{dU^\mu}{d\tau} = qF^{\mu\nu}U_\nu \tag{4.102}$$

The temporal component of the above equation is interesting,

$$\frac{\dot{E}}{c}\cosh\zeta = q\frac{\mathbf{v}\cdot\mathbf{E}}{c}\cosh\zeta \tag{4.103a}$$

$$\therefore \frac{dE}{dt} = q\mathbf{v}\cdot\mathbf{E} \tag{4.103b}$$

indicating the energy, E, imparted to the charged particle per unit time as it moves through an electric field, $\mathbf{E}$. Since the magnetic force is always perpendicular to the direction of motion, there is no way for a magnetic field to directly impart energy to a particle. Such energy is imparted instead by the counter-reaction of constraining forces that confine the particle's movement (e.g., in a wire).

If we replace the discrete charge quantity in (4.102) with a charge density, then we are left with a force density equation,

$$f^\mu = F^{\mu\nu}(\rho_0 U_\nu) = F^{\mu\nu}J_\nu \tag{4.104}$$

This force density can also be expressed in terms of the stress-energy tensor [14],

$$\partial_\nu T^{\mu\nu} = -\frac{1}{2\mu_0}\left(\partial_\nu\left(F^\mu{}_\sigma F^{\nu\sigma}\right) + \partial_\nu\left(G^\mu{}_\sigma G^{\nu\sigma}\right)\right) \tag{4.105a}$$

Table 4.4
Classical and Relativistic Formulation of Electromagnetic Laws

Name	Classical Equation(s)	Tensor Equation
Electromagnetic potential (Lorenz gauge)	$\Box^2\mathbf{A} = \mu_0\mathbf{J}$ $\Box^2\varphi = \rho/\varepsilon_0$	$\partial_\mu\partial^\mu A^\nu = \mu_0 J^\nu$
Lorenz gauge condition	$\boldsymbol{\nabla}\cdot\mathbf{A} + \frac{1}{c^2}\frac{\partial\varphi}{\partial t} = 0$	$\partial_\mu A^\mu = 0$
Electromagnetic fields	$\mathbf{E} = -\boldsymbol{\nabla}\varphi - \frac{\partial\mathbf{A}}{\partial t}$ $\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A}$	$F^{\mu\nu} = 2\partial^{[\mu}A^{\nu]}$
Maxwell's equations (inhomogeneous)	$\boldsymbol{\nabla}\cdot\mathbf{D} = \rho$ $\boldsymbol{\nabla}\times\mathbf{H} - \frac{\partial\mathbf{D}}{\partial t} = \mathbf{J}$	$\partial_\mu F^{\mu\nu} = \mu_0 J^\nu$
Maxwell's equations (homogeneous)	$\boldsymbol{\nabla}\cdot\mathbf{B} = 0$ $\boldsymbol{\nabla}\times\mathbf{E} + \frac{\partial\mathbf{B}}{\partial t} = 0$	$\partial_{[\mu}F_{\nu\sigma]} = 0$, or $\partial_\nu G^{\mu\nu} = 0$
Continuity equation	$\boldsymbol{\nabla}\cdot\mathbf{J} + \frac{\partial\rho}{\partial t} = 0$	$\partial_\mu J^\mu = 0$
Momentum and energy	$\boldsymbol{\sigma} = \mathbf{E}\otimes\mathbf{D} + \mathbf{B}\otimes\mathbf{H} - u\mathbf{I}$ $\mathbf{S} = \mathbf{E}\times\mathbf{H}$	$T^{\mu\nu} = \frac{-1}{2\mu_0}\left(F^\mu{}_\sigma F^{\nu\sigma} + G^\mu{}_\sigma G^{\nu\sigma}\right)$
Lorentz force law	$m\frac{d\mathbf{v}}{dt} = q\left(\mathbf{E} + \mathbf{v}\times\mathbf{B}\right)$	$m\frac{dU^\mu}{d\tau} = qF^{\mu\nu}U_\nu$, or $f^\mu = F^{\mu\nu}J_\nu = -\partial_\nu T^{\mu\nu}$

$$= -\frac{1}{\mu_0}\left(F^\mu{}_\sigma\left(\partial_\nu F^{\nu\sigma}\right) + G^\mu{}_\sigma\left(\partial_\nu G^{\nu\sigma}\right)\right) = -F^\mu{}_\sigma J^\sigma = -f^\mu + 0 \tag{4.105b}$$

$$\therefore f^\mu = -\partial_\nu T^{\mu\nu} \tag{4.105c}$$

A summary of the fundamental electromagnetic laws expressed in terms of spacetime tensors and four-vectors is given in Table 4.4.

4.6 COVARIANT WAVE EQUATIONS

We now return to the basic underlying phenomenon discovered by Maxwell that motivated the Lorentz transformations in the first place — the propagation of electromagnetic energy at the speed of light.

4.6.1 Wave Equations

Although I did not say so explicitly at the time, we have already seen a version of the wave equation in a Lorentz-covariant form, namely (4.46) for the four-potential. In a source-free region ($J^\nu = 0$), this becomes

$$\partial_\mu \partial^\mu A^\nu = 0 \tag{4.106}$$

where the left side is simply the d'Alembertian operator applied to the four-potential vector. We may similarly derive a wave equation for the electromagnetic field tensor itself. Let us start with the homogeneous Maxwell equation,

$$\partial_{[\mu} F_{\nu\sigma]} = \tfrac{1}{3} \left(\partial_\mu F_{\nu\sigma} + \partial_\sigma F_{\mu\nu} + \partial_\nu F_{\sigma\mu} \right) = 0 \tag{4.107a}$$

$$\therefore \partial_\sigma F_{\mu\nu} = -\partial_\nu F_{\sigma\mu} - \partial_\mu F_{\nu\sigma} \tag{4.107b}$$

and differentiate again using ∂^σ,

$$\partial_\sigma \partial^\sigma F_{\mu\nu} = -\partial_\nu \left(\partial^\sigma F_{\sigma\mu} \right) - \partial_\mu \left(\partial^\sigma F_{\nu\sigma} \right) = -\mu_0 \partial_\nu J_\mu + \mu_0 \partial_\mu J_\nu \tag{4.108}$$

where we have substituted for the parenthetical terms from the inhomogeneous Maxwell equation. Once again, if the region is source-free ($J_\mu = J_\nu = 0$), we have

$$\partial_\sigma \partial^\sigma F_{\mu\nu} = \partial_\sigma \partial^\sigma F^{\mu\nu} = 0 \tag{4.109}$$

which on the left is the d'Alembertian operator applied directly to the field tensor.

Example: We may verify that this is a wave equation by substituting into it a solution that has the form of a traveling wave — for example, propagating in the z direction with phase velocity v_p,

$$F^{\mu\nu} = F\left(z - v_p t\right) \tag{4.110a}$$

$$\therefore \partial_\sigma \partial^\sigma F\left(z - v_p t\right) = \left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2} \right) F\left(z - v_p t\right) \tag{4.110b}$$

$$= \frac{v_p^2}{c^2} F''\left(z - v_p t\right) - F''\left(z - v_p t\right) = 0 \tag{4.110c}$$

which is satisfied for all z and t when $v_p = c$. Thus, $F^{\mu\nu} = F(z - ct)$.

There are other constraints; the elements of the field tensor are not arbitrary. Aside from the prescribed z-dependence, they must also satisfy Maxwell's equations, which will limit, among other things, how the electric and magnetic components may relate to one another. For example, the inhomogeneous equation applied to a tensor with only the given $z - ct$ dependence (and no other spatial or temporal dependencies) in a source-free region is

$$\partial_\nu F^{\mu\nu} = \frac{1}{c}\begin{pmatrix} 0 & -\mathbf{E}^T \\ \mathbf{E} & \eta_0 \left[\mathbf{H}\right]_\times \end{pmatrix}\begin{pmatrix} \frac{1}{c}\frac{\partial}{\partial t} \\ \boldsymbol{\nabla} \end{pmatrix} = \frac{1}{c}\begin{pmatrix} -\boldsymbol{\nabla}\cdot\mathbf{E} \\ -\eta_0 \boldsymbol{\nabla}\times\mathbf{H} + \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} \end{pmatrix} \tag{4.111a}$$

$$= \frac{1}{c}\begin{pmatrix} -E'_z \\ \eta_0\left(H'_y\mathbf{x} - H'_x\mathbf{y}\right) - \mathbf{E}' \end{pmatrix} = 0 \tag{4.111b}$$

showing that the electric field (or at least its longitudinal slope, $\mathbf{E}'$) has no z component. Similarly, the homogeneous equation tells us that the longitudinal slope of the magnetic field likewise has no z-component,

$$\partial_\nu G^{\mu\nu} = \frac{1}{c}\begin{pmatrix} 0 & -\eta_0\mathbf{H}^T \\ \eta_0\mathbf{H} & -\left[\mathbf{E}\right]_\times \end{pmatrix}\begin{pmatrix} \frac{1}{c}\frac{\partial}{\partial t} \\ \boldsymbol{\nabla} \end{pmatrix} = \frac{1}{c}\begin{pmatrix} -\eta_0\boldsymbol{\nabla}\cdot\mathbf{H} \\ \boldsymbol{\nabla}\times\mathbf{E} + \frac{\eta_0}{c}\frac{\partial \mathbf{H}}{\partial t} \end{pmatrix} \tag{4.112a}$$

$$= \frac{1}{c}\begin{pmatrix} -\eta_0 H'_z \\ E'_x\mathbf{y} - E'_y\mathbf{x} - \eta_0\mathbf{H}' \end{pmatrix} = 0 \tag{4.112b}$$

Taken together, these equations show that we must have both the electric and magnetic fields perpendicular to the direction of propagation, and $\mathbf{E}' = \eta_0\mathbf{H}'\times\mathbf{z}$. In terms of the field tensor,

$$F'^{\mu\nu} = \frac{\eta_0}{c}\begin{pmatrix} 0 & -(\mathbf{H}'\times\mathbf{z})^T \\ \mathbf{H}'\times\mathbf{z} & \left[\mathbf{H}'\right]_\times \end{pmatrix} = \begin{pmatrix} 0 & -B'_y & B'_x & 0 \\ B'_y & 0 & 0 & B'_y \\ -B'_x & 0 & 0 & -B'_x \\ 0 & -B'_y & B'_x & 0 \end{pmatrix} \tag{4.113}$$

This particular type of solution is called a plane wave [15].

References

[1] Wikipedia. (2023, December) Ricci calculus. https://en.wikipedia.org/wiki/Ricci_calculus.

[2] G. Arfken and H. Weber, *Mathematical Methods for Physicists*, 4th ed. San Diego, CA: Academic Press, 1995.

[3] L. Susskind and A. Friedman, *Special Relativity and Classical Field Theory: The Theoretical Minimum*. New York: Basic Books, 2017.

[4] Wikipedia. (2019, September) Four-gradient. https://en.wikipedia.org/wiki/Four-gradient.

[5] Wikipedia. (2019, September) Gauge fixing. https://en.wikipedia.org/wiki/Gauge_fixing.

[6] J. M. Charap, *Covariant Electrodynamics: A Concise Guide*. Maryland: Johns Hopkins University Press, 2011.

[7] I. L. Shapiro, *A Primer in Tensor Analysis and Relativity*. Switzerland: Springer, 2019.

[8] Wikipedia. (2019, September) Electromagnetic tensor. https://en.wikipedia.org/wiki/Electromagnetic_tensor.

[9] Wikipedia. (2024, January) Hodge star operator. https://en.wikipedia.org/wiki/Hodge_star_operator.

[10] A. Einstein, "Zur elektrodynamik bewegter körper (on the electrodynamics of moving bodies)," *Annalen der Physik*, vol. 322, no. 10, pp. 891–921, June 1905.

[11] M. Schwartz, *Principles of Electrodynamics*. New York: Dover, 1987.

[12] Wikipedia. (2019, January) Electromagnetic stress-energy tensor. https://en.wikipedia.org/wiki/Electromagnetic_stress-energy_tensor.

[13] Wikipedia. (2023, October) Stress-energy tensor. https://en.wikipedia.org/wiki/Stress-energy_tensor.

[14] Wikipedia. (2023, December) Covariant formulation of classical electromagnetism. https://en.wikipedia.org/wiki/Covariant_formulation_of_classical_electromagnetism.

[15] C. Balanis, *Advanced Engineering Electromagnetics*. New York: Wiley, 1989.

Chapter 5

The Calculus of Spacetime

While tensor analysis using the format described in Chapter 4 has a rich history and is still very common, there are times when carrying the indices forward through complex calculations becomes rather cumbersome. To overcome this, we need a coordinate-free treatment of spacetime — one in which the physical significance and form of mathematical operators is not tied directly to the coordinate-system-specific representation of their operands (in the same way that the generic symbol $\boldsymbol{\nabla}$ replaces the direct Cartesian form $\mathbf{x}\partial_x + \mathbf{y}\partial_y + \mathbf{z}\partial_z$ in classical vector calculus). For special relativity, that compact symbolic representation is provided by an advanced mathematical formalism known as *spacetime algebra (STA)* [1–6], a special case of an even broader subject known as *Clifford algebra* or *geometric algebra* [7].

5.1 GEOMETRIC ALGEBRA

Geometric algebra extends the concept of a vector in multidimensional space (like the four-dimensional space of Minkowski spacetime) to higher-dimensioned counterparts such as oriented planar patches and volumes, and then provides the algebraic rules necessary to combine and manipulate those objects under a unified framework. This allows complex laws such as Maxwell's equations to be expressed and manipulated in a remarkably compact form.

5.1.1 Basis Vectors

First, in order to eliminate the indices attached to our tensor expressions, we construct any arbitrary four-vector, a, from a basis of orthogonal four-vectors, γ_μ

(not to be confused with the Lorentz factor, γ),

$$a = a^\mu \gamma_\mu = a^0\gamma_0 + a^1\gamma_1 + a^2\gamma_2 + a^3\gamma_3 \tag{5.1}$$

where, by convention, γ_0 is our timelike component and γ_1 to γ_3 are the spatial components. Normalization of the basis is given by the Minkowski metric,

$$\gamma_\mu \cdot \gamma_\nu = \eta_{\mu\nu} \tag{5.2}$$

Therefore, we have $\gamma_0^2 = +1$ and $\gamma_k^2 = -1$ for $k \geq 1$ in accordance with the $(+---)$ metric signature. Orthogonality is ensured since $\gamma_\mu \cdot \gamma_\nu = 0$ for $\mu \neq \nu$.

Although division of vectors is not supported per se, a multiplicative inverse, which is generally noncommutative, can usually be constructed having the form

$$a^{-1} = \frac{a}{a^2} \tag{5.3}$$

where a^2 is a scalar. Each standard basis vector, γ_μ, thus has a reciprocal, $\gamma^\mu = (\gamma_\mu)^{-1}$, defined such that the inner product is

$$\gamma_\mu \cdot \gamma^\nu = \delta_\mu{}^\nu = \begin{cases} 1 & \mu = \nu \\ 0 & \mu \neq \nu \end{cases} \tag{5.4}$$

where $\delta_\mu{}^\nu$ is a tensor representation of the Kronecker delta. These reciprocals may differ from their standard basis counterparts by at most a sign change in accordance with the metric signature; $\gamma^0 = \gamma_0$ and $\gamma^k = -\gamma_k$ for $k \geq 1$. The projection of vector a onto a unit vector in either the standard basis or the reciprocal basis is extracted by taking the inner product with the desired component,

$$a \cdot \gamma^\mu = a^\mu \tag{5.5a}$$

$$a \cdot \gamma_\nu = a_\nu \tag{5.5b}$$

Note that any four-vector can be converted to a new basis, say $\{e_\nu\}$, by first projecting onto the elements of the new reciprocal basis using the dot product $(a \cdot e^\nu)$ and then contracting the indices with the new standard basis vectors,

$$a' = (a \cdot e^\nu)\, e_\nu = a^\nu e_\nu = a^0 e_0 + a^1 e_1 + a^2 e_2 + a^3 e_3 \tag{5.6}$$

Changing the basis does not fundamentally change what the vector represents $(a' = a)$, therefore any four-vector, a, can be recast in terms of any desired basis as follows

$$a = (a \cdot e^\nu)\, e_\nu = (a \cdot e_\nu)\, e^\nu \tag{5.7}$$

5.1.2 Geometric Product

One of the most powerful concepts in geometric algebra is the *geometric product* or *Clifford product*, written simply as the juxtaposition, ab, for any two vectors, a and b, that are defined in the basis. To see what this means, we may expand the binomial form $(a + b)^2$, assuming that the geometric product is distributive over addition,

$$(a + b)^2 = a^2 + (ab + ba) + b^2 \tag{5.8a}$$

$$\therefore ab + ba = (a + b)^2 - a^2 - b^2 \tag{5.8b}$$

Note that we do not assume the geometric product is commutative (that is, in general $ab \neq ba$). We do assume that the sum of two vectors is another vector, and that the square of any vector (that is, the geometric product of any vector with itself) is a scalar. Thus, the right side of (5.8b), as the sum of squared vectors, must be a scalar. We use the inner product, or dot product, to fulfill this role,

$$\tfrac{1}{2}(ab + ba) = a \cdot b \tag{5.9}$$

This, then, is the symmetric part of the geometric product.

It remains to define the antisymmetric part. For that, we use the outer product, or *wedge product*, as follows,

$$\tfrac{1}{2}(ab - ba) = a \wedge b \tag{5.10}$$

The result of the wedge product is not a scalar, but a different kind of object known as a *bivector*. In the same sense that a vector represents a directed line segment, a bivector may be thought of as a directed plane segment, or an oriented planar patch [1, 2], as shown in Figure 5.1(a). The magnitude of the bivector represents the area subtended by the parallelogram formed by the two component vectors, and its orientation is in the sense of a sweeping toward b. Importantly, the shape of the patch is not encoded in the bivector, only its attitude (the plane in which it resides), its orientation (e.g., handedness), and its area. This is illustrated in Figure 5.1(b). As already described, the wedge product is antisymmetric, $a \wedge b = -b \wedge a$, illustrated in Figure 5.1(c). This further implies that the wedge product of any vector with itself (or with any other vector to which it is colinear) is identically zero.

In three-dimensional space, the cross product is closely related to the wedge product, but the latter is defined in such a way as to be extensible to higher dimensions, whereas the former is not. Also, in three dimensions the cross product

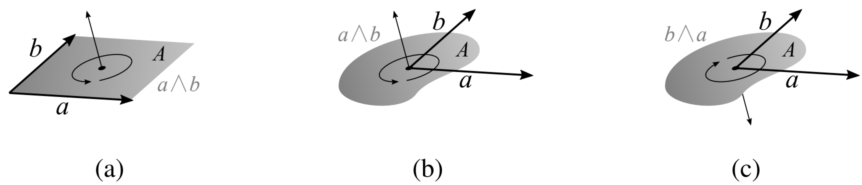


Figure 5.1 (a) Oriented planar patch of area A represented by the bivector $a \wedge b$. (b) Equivalent planar patch for $a \wedge b$. (c) Equivalent patch for $b \wedge a$.

of two vectors is another vector, but in higher dimensional space the wedge product of two vectors, the bivector, is a different kind of object entirely. One cannot define a unique normal vector to a two-dimensional plane in four dimensions [1].

The complete geometric product is then written as the sum of the symmetric and antisymmetric parts, or of the dot and wedge products, respectively,

$$ab = \tfrac{1}{2}(ab + ba) + \tfrac{1}{2}(ab - ba) = a \cdot b + a \wedge b \tag{5.11}$$

As stated previously, the geometric product is left and right distributive over addition,

$$a(b + c) = ab + ac \tag{5.12a}$$

$$(a + b)c = ac + bc \tag{5.12b}$$

It is also associative,

$$(ab)c = a(bc) \tag{5.13}$$

Conventionally, inner products take precedence over outer products, and those over geometric products in the absence of brackets. For example,

$$a \cdot bc = (a \cdot b)c \tag{5.14a}$$

$$ab \wedge c = a(b \wedge c) \tag{5.14b}$$

but I will often retain the unnecessary parentheses for clarity.

It should be noted that only the geometric product can be inverted in the manner shown in (5.3); the dot and wedge products alone do not embed enough information in their results to fully recover one operand given knowledge of the other [1]. For example, given the equation

$$ab = c \tag{5.15}$$

we can multiply both sides by b^{-1} to solve for a,

$$a = abb^{-1} = cb^{-1} = \frac{cb}{b^2} \tag{5.16}$$

However, there is nothing we can do with the following equation

$$a \cdot b = c \tag{5.17}$$

to divide out the dot product and solve for a; the equation simply does not provide enough information by itself. We must also furnish an expression for $a \wedge b$, which can then be combined with the dot product to form a full geometric product as in (5.11), and then solve for a using b^{-1} as indicated above.

5.1.3 Multivectors

One might reasonably question the meaning of an equation like (5.11) wherein a scalar ($a \cdot b$) is added to a bivector ($a \wedge b$). In the parlance of geometric algebra, scalars are called grade-0 objects, while vectors are grade-1, bivectors are grade-2, and so on. How can such things be added? For present purposes, we simply consider the two terms as representing different parts of a mixed quantity, just as real and imaginary parts are summed to create complex numbers, or $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ vector components are added to create a three-dimensional vector. Addition is simply a convenient way to group these quantities together so that they may be manipulated as one entity (called a *multivector*) using consistent and mutually accepted algebraic rules. Equations involving such mixed quantities are in fact systems of equations which can be separated (by grade, in this case, just like complex equations are separated into real and imaginary parts). Though it may seem awkward at first, the ability to mix different grades of objects under one label may rightly be considered geometric algebra's greatest strength, and is the key to it delivering remarkably compact, elegant expressions for otherwise rich and complex phenomena.

The orthogonality of the basis vectors ensures that their geometric products are the same as their wedge products, which are bivectors, unless both basis vectors are the same, in which case the metric signature gives the scalar result,

$$\gamma_\mu \gamma_\nu = \gamma_\mu \cdot \gamma_\nu + \gamma_\mu \wedge \gamma_\nu = \begin{cases} \eta_{\mu\mu} & \mu = \nu \\ \gamma_\mu \wedge \gamma_\nu & \mu \neq \nu \end{cases} \tag{5.18}$$

The bivectors resulting from the product of basis vectors are often written with a double-subscript for convenience,

$$\gamma_\mu \gamma_\nu = \gamma_\mu \wedge \gamma_\nu = \gamma_{\mu\nu} \quad (\mu \neq \nu) \tag{5.19}$$

We thus have six possible basis bivectors in four-dimensional Minkowski spacetime:

$$\{\gamma_{01}, \gamma_{02}, \gamma_{03}, \gamma_{12}, \gamma_{13}, \gamma_{23}\} \tag{5.20}$$

Products of the same vector pairs but with different ordering are equivalent to those above within a sign change (e.g., $\gamma_{12} = -\gamma_{21}$).

The geometric product of a vector and a bivector is a grade-3 object known, predictably, as a *trivector*. Trivectors represent oriented volumes in the same way that bivectors represent oriented plane segments, and vectors represent oriented line segments. In terms of spacetime algebra, there are four basis trivectors:

$$\{\gamma_{012}, \gamma_{013}, \gamma_{023}, \gamma_{123}\} \tag{5.21}$$

The lone grade-4 product of basis vectors in spacetime algebra, γ_{0123}, rather than being called a quadvector, is termed a pseudoscalar. Antisymmetry and the normalization of the basis vectors guarantee that the square of this object is -1,

$$\gamma_{0123}^2 = \gamma_0\gamma_1\gamma_2\gamma_3\gamma_0\gamma_1\gamma_2\gamma_3 = \gamma_0\gamma_1\gamma_2\gamma_3\gamma_3\gamma_2\gamma_1\gamma_0 \tag{5.22a}$$

$$= \left(\gamma_0^2\right)\left(\gamma_1^2\right)\left(\gamma_2^2\right)\left(\gamma_3^2\right) = (+1)(-1)(-1)(-1) = -1 \tag{5.22b}$$

The pseudoscalar is often shorthanded as i for this reason. The pseudoscalar commutes with even-grade elements (including itself), but anticommutes with odd-grade elements.

This completes the domain of the spacetime algebra. No higher-grade objects are needed, for any product of more than four vectors reduces to this basis set of 16 elements,

$$\{1\}, \{\gamma_\mu\}, \{\gamma_{\mu\nu}\}, \{\gamma_{\mu\nu\sigma}\}, \{i\} \tag{5.23}$$

(The first set element, $\{1\}$, is the basis element for scalar quantities.) It is conventional to write the first three basis bivectors (those containing γ_0) as

$$\sigma_k = \gamma_{k0} = \gamma_k\gamma_0 \quad k = 1, 2, 3 \tag{5.24}$$

It can be shown, then, that the remaining three basis bivectors are given, to within a sign change, by $i\sigma_k$. For example,

$$\gamma_{12} = \gamma_1\gamma_2 = -\gamma_0^2\gamma_1\gamma_2\gamma_3^2 = -\left(\gamma_0\gamma_1\gamma_2\gamma_3\right)\left(\gamma_3\gamma_0\right) = -i\sigma_3 \tag{5.25}$$

Furthermore, i itself can be expanded as $i = \sigma_1\sigma_2\sigma_3$. Similarly, the trivectors can all be written as $i\gamma_\mu$, and are sometimes called pseudovectors.

I may occasionally refer to the $i\sigma_k$ as *pseudobivectors* in order to distinguish them from σ_k, but strictly speaking all are merely bivectors of differing orientation in spacetime.[1] It is important to realize that this distinction is specific to a particular reference frame. The only difference between σ_k and $i\sigma_k$ is their orientation with respect to the coordinate system of a chosen worldline, γ_0, and that depends on the observer. In contrast, the difference between a vector and a pseudovector (between γ_μ and $i\gamma_\mu$) is unambiguously frame-independent; one is a directed line segment while the other is a three-volume, notions of dimensionality which have no relationship to the reference frame used.

We may thus alternately write the complete domain of spacetime algebra as

$$\{1\}, \{\gamma_\mu\}, \{\sigma_k, i\sigma_k\}, \{i\gamma_\mu\}, \{i\} \tag{5.26}$$

where

$$\mu = 0, 1, 2, 3 \tag{5.27a}$$

$$k = 1, 2, 3 \tag{5.27b}$$

We shall see that the basis components are interrelated by two distinct forms of duality, which are illustrated in Figure 5.2. Right multiplication by i^{-1} will be associated with Hodge duality, first described in Chapter 4. Right multiplication with a chosen timelike axis, or worldline, γ_0, describes *frame duality*, and will be associated with the connection between Minkowski spacetime and Euclidean space in a process called the *spacetime split*, to be described later in Section 5.1.8.

5.1.4 The Imaginary Unit

In complex analysis, we are taught that the imaginary unit, written as i in many contexts but more often as j in electrical engineering, is a quantity which squares to -1. No other characteristics are given, as if this property alone fully defines what it is. In geometric algebra, we have seen that a great many objects have this property,

- $\gamma_k^2 = -1 \quad$ for $k = 1, 2, 3$;
- $\gamma_{12}^2 = -1$;
- $\gamma_{012}^2 = -1$;

1 There is also a sense in which σ_k may be called polar bivectors, while $i\sigma_k$ are axial bivectors, as we will later see.

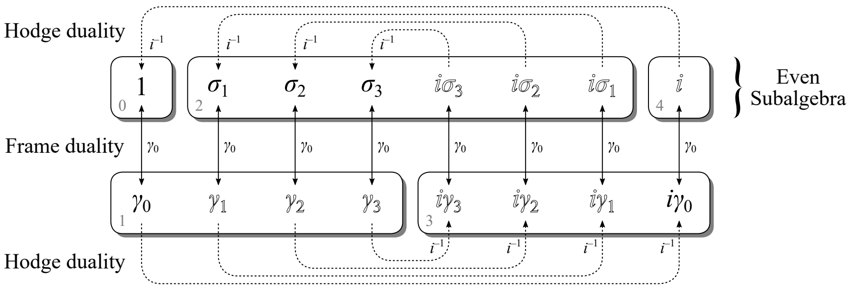


Figure 5.2 Two kinds of duality exhibited by the basis components of spacetime algebra. Components are separated by grade (indicated by the gray numeral in the lower left of each box). The Hodge dual is found by right multiplication with i^{-1}, and is reversed by right multiplication with i. Frame duality is found by right multiplication with a chosen timelike axis, or worldline, γ_0, and is bidirectional when the (+ − − −) metric is used. Solid symbols have a square norm of 1, while outlined symbols have a square norm of -1.

- $\gamma_{0123}^2 = -1$;

to name only a few. Should we then equate all of these objects to i, as though each is merely a different way of writing the same thing? Not at all. Both $+2$ and -2 square to 4, but that does not mean they are equal to one another. Each of the objects squaring to -1 above has a separate existence (not least of all because each is a different grade of object).

What is more important is that the sole property of squaring to -1, though it does not uniquely identify an object, is all that is required for it to fulfill the roles for which i is typically utilized in complex analysis. Most especially, for any object q which has this property, we may write a form of Euler's formula,

$$e^{q\theta} = \sum_{n=0}^{\infty} \frac{(q\theta)^n}{n!} = 1 + (q\theta) + \frac{(q\theta)^2}{2!} + \frac{(q\theta)^3}{3!} + \frac{(q\theta)^4}{4!} + \cdots \tag{5.28a}$$

$$= \left(1 + \frac{(q\theta)^2}{2!} + \frac{(q\theta)^4}{4!} + \cdots\right) + \left((q\theta) + \frac{(q\theta)^3}{3!} + \frac{(q\theta)^5}{5!} + \cdots\right) \tag{5.28b}$$

$$= \left(1 + q^2\frac{\theta^2}{2!} + q^4\frac{\theta^4}{4!} + \cdots\right) + q\left(\theta + q^2\frac{\theta^3}{3!} + q^4\frac{\theta^5}{5!} + \cdots\right) \tag{5.28c}$$

$$= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots\right) + q\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots\right) \tag{5.28d}$$

$$\therefore e^{q\theta} = \cos\theta + q\sin\theta \tag{5.28e}$$

Note that, for most of the objects listed earlier, this result is a mixed-grade, multivector quantity, where the $\cos\theta$ term is a scalar and $q\sin\theta$ has the same grade as q. This behavior will be exploited later as an operator for spatial rotations and Lorentz boosts, which take the form of rotations in Minkowski spacetime.

5.1.5 Grade Projection and Multivector Products

It is sometimes useful to isolate those components of a multivector, $\mathcal{M}$, corresponding to a certain grade. This is called *grade projection*, written $\langle\mathcal{M}\rangle_k$, representing the grade-k components of the multivector $\mathcal{M}$. Thus, for example,

$$\left\langle \pi + 2\gamma_2 + 3\sigma_1 + i\sqrt{2}\sigma_2 - i7\gamma_3 \right\rangle_2 = 3\sigma_1 + i\sqrt{2}\sigma_2 \tag{5.29}$$

It follows that any multivector is the sum of its grade projections,

$$\mathcal{M} = \sum_{k=0}^{n} \langle\mathcal{M}\rangle_k \tag{5.30}$$

The grade-0 (scalar) projection is sometimes written without the subscript, that is, $\langle\mathcal{M}\rangle_0 = \langle\mathcal{M}\rangle$.

Multivectors having only one nonzero grade, k, are said to be *homogeneous,* called *k-vectors*, and those which can be written as a single outer product of vectors are called *k-blades* [5]. The geometric product of a k-vector, K, and an l-vector, L, has the following decomposition,

$$KL = \langle KL\rangle_{k+l} + \langle KL\rangle_{k+l-2} + \cdots + \langle KL\rangle_{|k-l|} \tag{5.31}$$

We may generalize the dot and wedge products for all grades, defined earlier only for vectors, as the lowest-order and highest-order terms in this decomposition [2],

$$K \cdot L = \langle KL\rangle_{|k-l|} \tag{5.32a}$$

$$K \wedge L = \langle KL\rangle_{k+l} \tag{5.32b}$$

unless one of the multipliers is a scalar (grade 0) in which case there is only a single term in the summation (5.31). We apply the wedge product to this term; the dot product in that case is automatically zero (this preserves commutational symmetries and avoids double-counting redundant terms in certain equations).

When one element is a vector, a, (5.31) reduces to the original form in (5.11),

$$aK = \langle aK \rangle_{k-1} + \langle aK \rangle_{k+1} = a \cdot K + a \wedge K \tag{5.33}$$

Based on the symmetry and antisymmetry of the dot and wedge products for vectors, respectively, the generalized dot product above is symmetric for odd k, while the generalized wedge product is symmetric for even k. Thus,

$$a \cdot K = \langle aK \rangle_{k-1} = \tfrac{1}{2} \left(aK - (-1)^k Ka \right) \tag{5.34a}$$

$$a \wedge K = \langle aK \rangle_{k+1} = \tfrac{1}{2} \left(aK + (-1)^k Ka \right) \tag{5.34b}$$

5.1.6 Interpretation of Products

Note that the generalized dot product of a vector and a k-vector given in (5.34a) does not always yield a scalar; rather, it decrements the grade of the operand, K, by one, and may be considered the same as tensor contraction in Ricci calculus. The wedge product, in contrast, increments the grade by one, and imitates tensor products having distinct, uncontracted indices [5].

The dot product effectively measures the colinearity, or alignment, between two vectors, and the classic magnitude-angle formula for the vector dot product holds in a form modified for spacetime,

$$a \cdot b = \sqrt{\left(a^\dagger \cdot a\right)\left(b^\dagger \cdot b\right)} \cos\left(\theta_{a^\dagger b}\right) \tag{5.35}$$

where the notation $a^\dagger$ indicates the *relative reversion* of vector a. More details about this operation are given in Section D.5.6.[2] For vectors, the relative reversion is given by

$$a^\dagger = \gamma_0 a \gamma_0 \tag{5.36}$$

and amounts in the case of metric signature $(+---)$ to the negation of the spatial components. Note that the angle $\theta_{a^\dagger b}$ is measured between $a^\dagger$ and b (or by symmetry between a and $b^\dagger$), as shown in Figure 5.3(a). Furthermore, if the metric were

2 While the essentials of spacetime algebra will be described in the main text as they are needed, Appendix D is intended as a more comprehensive reference with emphasis on rigor and completeness. Specific useful results such as mathematical identities, which would be too tedious and distracting to derive in the main text, shall be pulled from this appendix when required. It is strongly suggested that readers who are new to spacetime algebra familiarize themselves with Appendix D, as it will be referred to often.

Figure 5.3 (a) The dot product measures the magnitude and colinearity between two vectors. (b) Dot products in Minkowski space must account for the metric signature; two vectors can be orthogonal in Minkowski space without appearing to lie at right angles in Euclidean space.

Euclidean, we would find that $a^\dagger \cdot a = |a|^2$ and the more familiar magnitude-angle form would hold. In Minkowski spacetime, this conversion corrects for any spatial distortions introduced by Lorentz transformation (i.e., the results are frame-independent).

Example: As a consequence, two vectors can be orthogonal in spacetime without appearing to lie at right angles to one another in the Euclidean representation of a particular reference frame, as illustrated in Figure 5.3(b). Take the following two vectors,

$$a = 2\gamma_0 + \gamma_1 \tag{5.37a}$$

$$b = \gamma_0 + 2\gamma_1 \tag{5.37b}$$

Their squares are

$$a^2 = 4\gamma_0^2 + \gamma_1^2 = 4 - 1 = +3 \tag{5.38a}$$

$$b^2 = \gamma_0^2 + 4\gamma_1^2 = 1 - 4 = -3 \tag{5.38b}$$

indicating that a is a timelike vector, while b is spacelike. Both have a magnitude of $\sqrt{3}$, but the dot product shows them to be orthogonal,

$$(2\gamma_0 + \gamma_1) \cdot (\gamma_0 + 2\gamma_1) = 2\left(\gamma_0^2 + \gamma_1^2\right) = 2(1 - 1) = 0 \tag{5.39}$$

suggesting that $\theta = \pi/2$. These vectors are plotted in Figure 5.3(b). Whereas a represents the worldline for a particular reference frame, b represents its associated surface of simultaneity. These are in fact orthogonal directions in the context of special relativity.

Wedge products with vectors are interpreted as sweeping a lower-dimensional component into a higher-level dimension. Thus, the wedge product of a vector and a

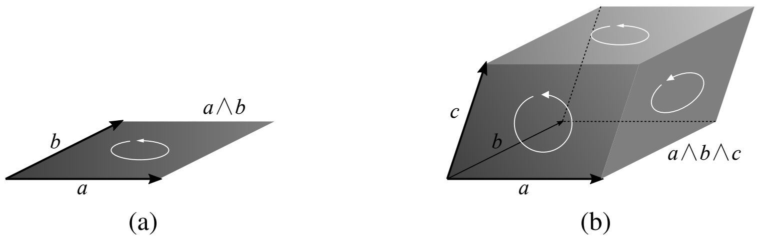


Figure 5.4 Wedge products extend blades into higher dimensions. (a) A vector is swept out into a bivector plane. (b) A bivector plane is extended up into a directed volume.

vector yields a bivector plane segment, while the product of a vector and a bivector plane yields a volume element, and so on, as shown in Figure 5.4. In all cases, the higher-dimensional result inherits its orientation parameters from the order of the elements in the wedge product.

The dot product of vectors is most often associated with *geometric projection*. To see how, we may expand the vector a in terms of its dot and wedge products with the vector b,

$$a = a\left(bb^{-1}\right) = (ab)b^{-1} = (a \cdot b)b^{-1} + (a \wedge b)b^{-1} \tag{5.40}$$

where the inverse of b is interpreted as given in (5.3),

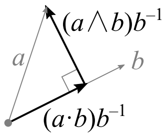
$$b^{-1} = \frac{b}{b^2} = \frac{b}{b \cdot b} \tag{5.41}$$

The first term in (5.40) is essentially identical to the classic vector projection formula,

$$(a \cdot b)b^{-1} = \frac{a \cdot b}{b \cdot b}b = \text{proj}_b a \tag{5.42}$$

and is effectively a change of basis, as anticipated in (5.7). The second term in (5.40) is the residual of a after its projection onto b is removed, and is known as the rejection of vector a from b. The projected and rejected vectors are shown in Figure 5.5(a).

The same expansion works for k-vectors and blades as well. The projection and rejection of a vector a onto and from a bivector plane K is shown in Figure 5.5(b). The projected vector, $(a \cdot K)K^{-1}$, lies in the plane of K, while the rejected vector, $(a \wedge K)K^{-1}$, is orthogonal to it.



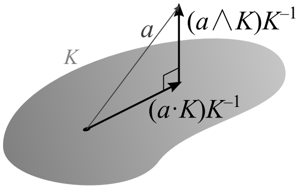


Figure 5.5 Visualization of the dot and wedge products as projection and rejection operators, respectively. (a) Projection and rejection of one vector onto and from another. (b) Projection and rejection of a vector onto and from a bivector plane. Perpendicularity in these figures is to be taken in the context of Minkowski spacetime — the indicated vectors may not actually appear perpendicular in a Euclidean representation, as described previously.

For convenience, we will sometimes adopt a subscript notation for projection operations as follows,

$$a_K = (a \cdot K)K^{-1} \tag{5.43}$$

and rejection will be written with a struck-out subscript,

$$a_{\not K} = (a \wedge K)K^{-1} \tag{5.44}$$

It can additionally be shown that the rejection of vector a from a k-vector, K, is equivalent to the projection onto the dual k-vector, iK,

$$a_{\not K} = a_{iK} \tag{5.45}$$

5.1.7 Simple Bivectors and the Canonical Form

Although the wedge product of two vectors is always a bivector, not all bivectors can be decomposed into the product of two vectors [8]. For example,

$$F = \sigma_1 + i\sigma_1 = \gamma_1 \wedge \gamma_0 + \gamma_3 \wedge \gamma_2 \tag{5.46}$$

cannot be written as the wedge product of just two vectors. This is because the two wedge products above involve four different spacetime axes. If there was an axis in common to both terms, it would be possible to factor that one out, leaving the other two summed in accordance with the distributive property (and the sum of two vectors is a vector), but four-dimensional spacetime has given us enough degrees of freedom to have two bivector planes which are neither intersecting nor parallel. Their summation is therefore irreducible. F is therefore a homogeneous 2-vector, but not a 2-blade.

One of the consequences of this is that the square norm of a bivector may not be a scalar. Consider the following expansion, which is generally true for all k-vectors,[3]

$$K^2 = K \cdot K + K \wedge K \tag{5.47}$$

For all k-vectors except for bivectors, the wedge product above is zero, leaving the scalar dot product as the result. But if a bivector composes the irreducible sum of two planes, as described above, then the wedge product between them will produce a four-volume (pseudoscalar) element. We thus say that the square norm of a bivector is, in general, complex.

Bivectors that can be written as a single wedge product are called simple bivectors. We can test for simplicity by verifying that the wedge product of the bivector with itself vanishes. Take, for example, the bivector σ_k,

$$\sigma_k \wedge \sigma_k = \gamma_{k0} \wedge \gamma_{k0} = \gamma_k \wedge \gamma_0 \wedge \gamma_k \wedge \gamma_0 = 0 \tag{5.48}$$

Repetition of any vector in a wedge product makes the result zero, so the appearance of γ_k and γ_0 twice in the product above ensures that σ_k is a simple bivector. In fact, all bivectors of the form $\mathbf{a} = a^k\sigma_k$ or $i\mathbf{b} = ib^k\sigma_k$ are simple bivectors, so it follows that any bivector can be written as the sum of at most two simple bivectors,

$$F = \mathbf{a} + i\mathbf{b} \tag{5.49}$$

Let us take the wedge product of F with itself to see under what conditions the combination may also be simple,

$$F \wedge F = (\mathbf{a} + i\mathbf{b}) \wedge (\mathbf{a} + i\mathbf{b}) \tag{5.50a}$$

$$= \mathbf{a} \wedge \mathbf{a} + (i\mathbf{b}) \wedge \mathbf{a} + \mathbf{a} \wedge (i\mathbf{b}) + (i\mathbf{b}) \wedge (i\mathbf{b}) \tag{5.50b}$$

Since we know that $\mathbf{a}$ and $i\mathbf{b}$ are simple, their self-products above are zero and can be dropped. The remaining terms can be simplified using an identity that exploits

3 One must be careful when generalizing results like this. What is true for k-vectors and blades may not be true for multivectors in general. And while $K^2 = K \cdot K + K \wedge K$ for all k-vectors, it is generally not the case that $KL = K \cdot L + K \wedge L$ unless one of K or L is grade 1. An extensive list of valid product identities is given in Appendix D.

Hodge duality between the dot and wedge products,

$$\mathbf{a} \wedge (i\mathbf{b}) = (\mathbf{a} \cdot \mathbf{b})i \tag{5.51}$$

It is also worth recalling immediately that i commutes with all even-grade objects, including $\mathbf{a}$, $\mathbf{b}$, the scalar that is their dot product, and the pseudoscalar that is their wedge product. Therefore,

$$F \wedge F = (i\mathbf{b}) \wedge \mathbf{a} + \mathbf{a} \wedge (i\mathbf{b}) = i(\mathbf{b} \cdot \mathbf{a}) + (\mathbf{a} \cdot \mathbf{b})i = 2(\mathbf{a} \cdot \mathbf{b})i \tag{5.52}$$

Thus, F is simple if and only if $\mathbf{a}$ and $\mathbf{b}$ are orthogonal ($\mathbf{a} \cdot \mathbf{b} = 0$), and is nonsimple otherwise. I will use boldface roman characters throughout this book to signify bivectors that are manifestly simple, like $\mathbf{a}$ and $\mathbf{b}$.

The problem with the decomposition in (5.49) is that it is reference-frame dependent, as described in Section 5.1.3. The bivector planes σ_k and $i\sigma_k$ can only be distinguished by their orientation with respect to γ_0, the chosen worldine of a particular observer. A representation that is not frame-dependent is given by the canonical polar form,

$$F = \mathbf{f}e^{i\varphi} = \mathbf{f}(\cos\varphi + i\sin\varphi) \tag{5.53}$$

where $\mathbf{f}$ is a simple bivector and φ is a scalar, which can be found by

$$\varphi = \frac{1}{2}\tan^{-1}\left(\frac{\sin(2\varphi)}{\cos(2\varphi)}\right) = \frac{1}{2}\tan^{-1}\left(\frac{\langle F^2\rangle_4}{i\,\langle F^2\rangle_0}\right) \tag{5.54a}$$

$$\therefore \mathbf{f} = Fe^{-i\varphi} \tag{5.54b}$$

(Computer codes should use a four-quadrant implementation of the arctangent function in the above formulas.)

If F is null ($F^2 = 0$), then the scalar phase φ is indeterminate (much like the phase of $z = 0 + i0$ in complex analysis). In that event, the decomposition above will not be unique; any phase at all can be selected, and then $\mathbf{f}$ found via (5.54b). Note that $\mathbf{f}$ will then be a null bivector like F, which by definition is also a simple bivector.

If F is non-null, then $\mathbf{f}$ is both simple and uniquely determined (within a sign; we could always substitute $\varphi' = \varphi + \pi$ and then negate $\mathbf{f}$, for the arctangent above would be unchanged).

Either way, we have ensured that $\mathbf{f}$ is simple, having a scalar square norm that is equal in magnitude to the complex square norm of its parent bivector,

$$\mathbf{f}^2 = |F|^2 \tag{5.55}$$

More importantly, the polar decomposition is frame-independent, allowing us to perform calculations on bivectors without immediately committing to a particular reference frame. For example, complex conjugation is defined for bivectors as a negation of the phase argument $i\varphi$ in the canonical form (5.53), not the negation of the pseudobivector $i\mathbf{b}$ in the frame-dependent Cartesian form (5.49),

$$F^* = \mathbf{f}e^{-i\varphi} \neq \mathbf{a} - i\mathbf{b} \tag{5.56}$$

This allows us to write a general formula for the multiplicative inverse,

$$F^{-1} = \frac{F^*}{F^*F} \tag{5.57}$$

If $F^*F = 0$ despite F having nonzero values, then F is null, and cannot be inverted. The same formula applies to any k-vector, with conjugation of other grades working as one would expect: pseudovectors ($i\gamma_\mu$) and pseudoscalars ($i\beta$) are negated, while all other grades are unaffected.

The ability to perform derivations with spacetime quantities in a way that is not dependent on reference frame is critical, but when a frame-specific description is desired — for example, when a problem has been solved and we wish to see how it will manifest itself to a particular observer — we can use the spacetime split, described in the next section.

5.1.8 Spacetime Split

Just as three-vectors represent absolute magnitude and direction in three-dimensional space, regardless of what coordinate system they are expressed in (e.g., Cartesian, cylindrical, spherical), spacetime four-vectors represent a magnitude and direction in spacetime that is independent of the reference frame implied by a particular set of basis vectors. To explicitly describe a four-vector quantity in a particular reference frame, we must do two things:

- First, project that four-vector onto a particular timelike axis, γ_0 (representing the worldline of the reference frame in question), yielding a scalar time-dependent component. This is accomplished by taking the dot product, $a \cdot \gamma_0$.

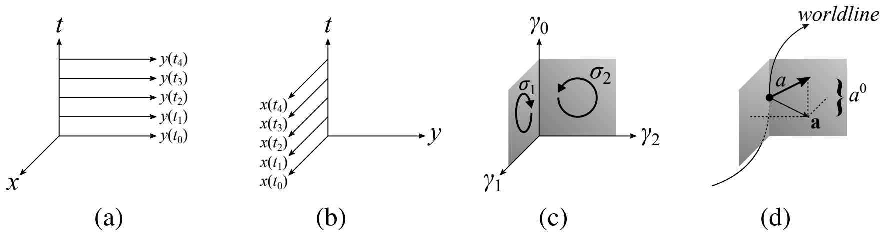


Figure 5.6 Illustration that a coordinate axis in three-dimensional space fixed in time is associated with a bivector plane in four-dimensional spacetime. (a) Fixed y-coordinate axes at multiple times, t. (b) Fixed x-coordinate axes at multiple times, t. (c) Bivector planes associated with those coordinate axes. (d) Spacetime split of a four-vector a into a spatial three-vector $\mathbf{a}$ and temporal offset a^0.

- Second, reject the temporal axis, yielding a three-dimensional space of vector components. This is most naturally accomplished using the wedge product, $a \wedge \gamma_0$.

The geometric product simultaneously accomplishes both of these things,

$$a\gamma_0 = \left(a^0\gamma_0 + a^k\gamma_k\right)\gamma_0 = a^0 + a^k\sigma_k = a^0 + \mathbf{a} \tag{5.58}$$

Note that we have identified the spacetime bivector $a^k\sigma_k$ with the spatial three-vector $\mathbf{a}$, represented in bold type because it is also a simple bivector. The four-dimensional bivectors, σ_k, may be associated with the coordinate axes of three-dimensional space (e.g., $\sigma_1 = \mathbf{x}$, $\sigma_2 = \mathbf{y}$, and $\sigma_3 = \mathbf{z}$). This is illustrated in Figure 5.6. Discarding the z spatial axis for the sake of illustration, we show the y axis in Figure 5.6(a) at multiple times, t_0 through t_4. Similarly, in Figure 5.6(b), we show the x axis vectors at multiple times. Converting to spacetime basis vectors in Figure 5.6(c), we highlight the bivector planes, σ_1 and σ_2, associated with these time-independent coordinate axes.

Thus, with a process referred to as the spacetime split, we have essentially converted a four-vector quantity, a, into a multivector, $a^0 + \mathbf{a}$, having a scalar (temporal) part and a vector (spatial) part, sometimes called a *paravector* [1], illustrated in Figure 5.6(d). Paravectors and spacetime four-vectors may be thought of as dual representations of a physical quantity — one which is frame-dependent and the other which is not. Transformation between the two (in either direction) is accomplished by right-multiplication with the worldline, γ_0, of the reference frame selected for paravector representation.

Table 5.1
Spacetime Split of Common Four-Vectors

Name	Symbol	Four-Vector Components				Spacetime Split	
		$\mu = 0$	$\mu = 1$	$\mu = 2$	$\mu = 3$	Expression	Paravector
Position	x	ct	x_1	x_2	x_3	$x\gamma_0$	$ct + \mathbf{r}$
Velocity*	U	$c\gamma$	$v_1\gamma$	$v_2\gamma$	$v_3\gamma$	$U\gamma_0$	$\gamma\,(c + \mathbf{v})$
Momentum	P	E/c	p_1	p_2	p_3	$P\gamma_0$	$E/c + \mathbf{p}$
Force*	F	$\dot{E}\gamma/c$	$\dot{p}_1\gamma$	$\dot{p}_2\gamma$	$\dot{p}_3\gamma$	$F\gamma_0$	$\gamma\left(\dot{E}/c + \mathbf{p}\right)$
Wavevector	K	ω/c	k_1	k_2	k_3	$K\gamma_0$	$\omega/c + \mathbf{k}$
Source density	J	$c\rho$	J_1	J_2	J_3	$J\gamma_0$	$c\rho + \mathbf{J}$
Potential	A	φ/c	A_1	A_2	A_3	$A\gamma_0$	$\varphi/c + \mathbf{A}$
Gradient	$\square$	$c^{-1}\partial_t$	∂_1	∂_2	∂_3	$\gamma^0\square$	$\partial_0 + \boldsymbol{\nabla}$

$^*\gamma = \cosh\zeta$.

Left-multiplication with γ_0 produces a paravector split that negates the spatial components,

$$\gamma_0 a = a^0 - \mathbf{a} \tag{5.59}$$

Therefore, we can calculate the square norm of any vector

$$a^2 = a\left(\gamma_0^2\right)a = (a\gamma_0)(\gamma_0 a) = \left(a^0 - \mathbf{a}\right)\left(a^0 + \mathbf{a}\right) = \left(a^0\right)^2 - \mathbf{a}^2 \tag{5.60}$$

While γ_0 made a brief appearance in above derivation, it is absent from the final result, indicating that the square norm is independent of the chosen worldine, γ_0. This very simply demonstrates a point that we articulated previously, namely that the square norm of a four-vector is invariant under Lorentz transformation [1, 6].

Each of the four-vectors in Table 4.1 has a corresponding spacetime split, as shown in Table 5.1.[4] By projecting spacetime quantities having four dimensions into the more familiar three-dimensional reference frame, the spacetime split puts the equations we will derive into a form that could be more easily recognized by one trained in classical vector calculus.

An important expansion in this context is the geometric product of two spatial three-vectors, which are simple bivectors in the σ_k basis (the imaginary unit, i, can

4 At this point, we begin using a lowercase x, representing the spacetime four-position, as the free independent variable for most problems. Other four-vectors will retain the capital-letter convention we established when they were introduced as tensor quantities, and boldface $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$, all simple bivectors, shall still refer to the three Cartesian axes in Euclidean space.

always be moved out of such a product owing to the associativity of the geometric product and its commutivity with even-grade components),

$$\mathbf{ab} = \left(a^1\sigma_1 + a^2\sigma_2 + a^3\sigma_3\right)\left(b^1\sigma_1 + b^2\sigma_2 + b^3\sigma_3\right) \tag{5.61a}$$

$$= \left(a^1b^1 + a^2b^2 + a^3b^3\right) + \sigma_2\sigma_3\left(a^2b^3 - a^3b^2\right) + \sigma_3\sigma_1\left(a^3b^1 - a^1b^3\right) + \sigma_1\sigma_2\left(a^1b^2 - a^2b^1\right) \tag{5.61b}$$

$$= \left(a^1b^1 + a^2b^2 + a^3b^3\right) + i\sigma_1\left(a^2b^3 - a^3b^2\right) + i\sigma_2\left(a^3b^1 - a^1b^3\right) + i\sigma_3\left(a^1b^2 - a^2b^1\right) \tag{5.61c}$$

$$= \mathbf{a}\cdot\mathbf{b} + i(\mathbf{a}\times\mathbf{b}) \tag{5.61d}$$

where the dot and cross products in equations such as these have their conventional meanings from three-dimensional vector calculus. Symmetry, then, allows us to isolate the dot and cross product forms,

$$\mathbf{a}\cdot\mathbf{b} = \tfrac{1}{2}(\mathbf{ab} + \mathbf{ba}) \tag{5.62a}$$

$$\mathbf{a}\times\mathbf{b} = -\tfrac{i}{2}(\mathbf{ab} - \mathbf{ba}) \tag{5.62b}$$

5.1.9 Differentiation

Let us now define the spacetime gradient operator as follows,

$$\Box = \gamma^\mu\partial_\mu = \gamma^0\partial_0 + \gamma^1\partial_1 + \gamma^2\partial_2 + \gamma^3\partial_3 \tag{5.63}$$

Here, we use the four-sided box symbol, $\Box$, to represent the four dimensions of spacetime over which the derivatives are taken. It is a simple matter to show that $\Box^2$ is the d'Alembertian operator given in (1.34). The three-dimensional gradient, $\boldsymbol{\nabla}$, is a component of the spacetime split of this four-dimensional spacetime gradient,

$$\gamma^0\Box = \partial_0 + \boldsymbol{\nabla} \tag{5.64}$$

The spacetime gradient has the same form and thus shares the same properties and expansions as other four-vectors in many respects, with the sole caveat that derivatives of products must properly take into account the product rule for differentiation, for example,

$$\Box(AB) = \dot{\Box}\dot{A}B + \dot{\Box}A\dot{B} \tag{5.65}$$

Overdots are used to indicate which terms in a product a particular differential operator applies to, when a clearer reordering of terms is precluded by commutivity constraints.

Finally, differentiation can be restricted to a *tangential derivative* along a particular k-vector or blade, K, by projection,

$$\Box_K = (\Box \cdot K)K^{-1} \tag{5.66}$$

While we have touched on the most critical aspects here in the main text, additional properties, identities, and detailed information about spacetime algebra are given in Appendix D.

5.2 ELECTROMAGNETIC LAWS IN SPACETIME ALGEBRA

I stated at the beginning of this chapter that geometric algebra would allow us to write Maxwell's equations in a remarkably compact form. We are now prepared to make good on that promise by translating those equations into the efficient algebra of spacetime. Even the most seasoned experts in electromagnetic theory, I predict, may be surprised by how simple the laws of their profession really are, given the proper mathematical framework.[5]

5.2.1 Inner Products of Four-Vectors

Recall that the square norm of any valid four-vector is a constant, invariant under Lorentz transformation. In tensor analysis, this norm is calculated by contracting the indices of the covariant and contravariant forms of that four-vector. For example, the proper time, τ, is derived from the four-position as follows,

$$(c\tau)^2 = X_\mu X^\mu \tag{5.67}$$

5 Claude Shannon, father of information theory and a personal inspiration of mine, said that "Much of the power and elegance of any mathematical theory. . . depends on use of a suitably compact and suggestive notation, which nevertheless completely describes the concepts involved" [9].

In spacetime algebra, the norm is simply given by the square of the index-free four-vector,

$$(c\tau)^2 = x^2 \tag{5.68}$$

Similarly, the norm of the four-velocity, given in tensor form by (4.20), may be written

$$U_\mu U^\mu = c^2 \quad \Longleftrightarrow \quad U^2 = c^2 \tag{5.69}$$

More generally, the inner product of two distinct four-vectors, once again taking the form of index contraction in tensor analysis, is associated with the dot product in spacetime algebra. For example, the Lorentz-invariant phase of a wave may be written in the following two equivalent forms,

$$K_\mu X^\mu = \Phi \quad \Longleftrightarrow \quad K \cdot x = \Phi \tag{5.70}$$

We may also write the continuity equation,

$$\partial_\mu J^\mu = 0 \quad \Longleftrightarrow \quad \Box \cdot J = 0 \tag{5.71}$$

For any multivector, $\mathcal{M}$, I will refer to $\Box \cdot \mathcal{M}$ as the *spacetime divergence* to distinguish it from the classical, three-vector divergence, $\boldsymbol{\nabla} \cdot \mathbf{a}$. Likewise, $\Box \wedge \mathcal{M}$ shall be called the *spacetime curl*.

5.2.2 Four-Potential and the Lorenz Gauge Condition

We have already stated that the square norm of the spacetime gradient operator is given by the d'Alembertian. This makes it easy to write the four-potential in terms of the four-current density,

$$\partial_\mu \partial^\mu A^\nu = \mu_0 J^\nu \quad \Longleftrightarrow \quad \Box^2 A = \mu_0 J \tag{5.72}$$

where we have assumed the Lorenz gauge condition. The Lorenz gauge condition itself translates quite simply,

$$\partial_\mu A^\mu = 0 \quad \Longleftrightarrow \quad \Box \cdot A = 0 \tag{5.73}$$

5.2.3 The Faraday Bivector Field

We found previously that the electromagnetic field is represented in special relativity not by four-vectors, but by a second-rank, antisymmetric tensor, $F^{\mu\nu}$. The equivalent object in spacetime algebra is a bivector.

Recall that the Faraday tensor was given by the antisymmetrized four-gradient of the four-potential,

$$F^{\mu\nu} = 2\partial^{[\mu} A^{\nu]} = \partial^\mu A^\nu - \partial^\nu A^\mu \tag{5.74}$$

In spacetime algebra, antisymmetrization is achieved via the wedge product,

$$F = \Box \wedge A \tag{5.75}$$

This translation from a second-rank tensor to a bivector requires some consideration; we have omitted the prefactor 2, and have not converted the spacetime gradient to the upper-index form shown in (5.74).[6] We can expand terms to establish an equivalence between the Faraday bivector field and its second-rank tensor,

$$F = \Box \wedge A = (\gamma^\mu \partial_\mu) \wedge (A^\nu \gamma_\nu) = (\gamma_\mu \partial^\mu) \wedge (A^\nu \gamma_\nu) \tag{5.76a}$$

$$= \partial^\mu A^\nu (\gamma_\mu \wedge \gamma_\nu) = \tfrac{1}{2}\left[\partial^\mu A^\nu (\gamma_\mu \wedge \gamma_\nu) + \partial^\nu A^\mu (\gamma_\nu \wedge \gamma_\mu)\right] \tag{5.76b}$$

$$= \tfrac{1}{2}(\partial^\mu A^\nu - \partial^\nu A^\mu)(\gamma_\mu \wedge \gamma_\nu) = \tfrac{1}{2} F^{\mu\nu} (\gamma_\mu \wedge \gamma_\nu) \tag{5.76c}$$

Alternatively, we could use a spacetime split to expand (5.75) into its paravector form,

$$F = \Box \wedge A = \tfrac{1}{2}\left(\Box A - \dot{A}\dot{\Box}\right) = \tfrac{1}{2}\left(\Box \gamma^0 \gamma_0 A - \dot{A}\gamma_0 \gamma^0 \dot{\Box}\right) \tag{5.77a}$$

$$= \tfrac{1}{2}(\partial_0 - \boldsymbol{\nabla})\left(A^0 - \mathbf{A}\right) - \tfrac{1}{2}\left(\dot{A}^0 + \dot{\mathbf{A}}\right)\left(\dot{\partial}_0 + \dot{\boldsymbol{\nabla}}\right) \tag{5.77b}$$

$$= -\partial_0 \mathbf{A} - \boldsymbol{\nabla} A^0 + \tfrac{1}{2}\left(\boldsymbol{\nabla}\mathbf{A} - \dot{\mathbf{A}}\dot{\boldsymbol{\nabla}}\right) \tag{5.77c}$$

6 More accurately, no conversion analogous to raising or lowering of indices is needed in spacetime algebra. The tedious index manipulation required to use Ricci calculus is an artifact of the way in which its forms are tied to a particular coordinate system and reference frame. Spacetime algebra is coordinate-free by design, first revealing and then exploiting the fact that, say, $\Box = \gamma^\mu \partial_\mu$ and $\Box = \gamma_\mu \partial^\mu$ are exactly the same object.

$$= -\frac{1}{c}\left(\frac{\partial \mathbf{A}}{\partial t} + \boldsymbol{\nabla}\varphi\right) + i\boldsymbol{\nabla} \times \mathbf{A} = \tfrac{1}{c}\mathbf{E} + i\mathbf{B} \tag{5.77d}$$

Right multiplication by $i^{-1} = -i$ produces the Hodge dual [1] of F,

$$Fi^{-1} = -Fi = \mathbf{B} - i\tfrac{1}{c}\mathbf{E} = G = \star F \tag{5.78}$$

It is no accident that $\mathbf{E}$, a polar vector as described in Section 2.1.3, is the real part of the field F, while $\mathbf{B}$, an axial vector, is the imaginary part. To put it another way, $\mathbf{E}$ is timelike, having its basis in the bivector planes involving time (γ_{10}, γ_{20}, γ_{30}) while $\mathbf{B}$ is spacelike, having its basis in the purely spatial bivector planes (γ_{12}, γ_{13}, γ_{23}). The geometric interpretation of what it means to be a polar or an axial vector is thus made explicit in spacetime algebra.

It is also interesting at this point to consider the square of the field bivector,

$$F^2 = \left(\tfrac{1}{c}\mathbf{E} + i\mathbf{B}\right)\left(\tfrac{1}{c}\mathbf{E} + i\mathbf{B}\right) = \left(\tfrac{1}{c^2}|\mathbf{E}|^2 - |\mathbf{B}|^2\right) + \tfrac{2i}{c}\left(\mathbf{E}\cdot\mathbf{B}\right) \tag{5.79}$$

Both terms above match, within a constant factor or sign, the field tensor invariants identified in Section 4.3.3. Moreover, the term that was previously identified as a pseudoscalar $(\mathbf{E}\cdot\mathbf{B})$ is a coefficient of i in this expression.

Note that if the Lorenz gauge is assumed ($\Box \cdot A = 0$), then the field bivector may be written directly in terms of the spacetime gradient,

$$F = \Box \cdot A + \Box \wedge A = \Box A \tag{5.80}$$

5.2.4 Maxwell's Equation

To write Maxwell's equations in spacetime algebra, we consider first the spacetime divergence of F,

$$\Box \cdot F = \Box \cdot (\Box \wedge A) = \Box^2 A - \Box(\Box \cdot A) \tag{5.81}$$

where we have used the identities (D.164) and (D.166). The last term in parentheses vanishes via the Lorenz gauge, leaving

$$\Box \cdot F = \Box^2 A = \mu_0 J \tag{5.82}$$

Table 5.2
Maxwell's Equations and Related Formulas in Various Forms

Name	Classical Equation(s)	Tensor Equation	Spacetime Algebra
Lorenz gauge condition	$\boldsymbol{\nabla}\cdot\mathbf{A}+\frac{1}{c^2}\frac{\partial\varphi}{\partial t}=0$	$\partial_\mu A^\mu=0$	$\square\cdot A=0$
Electromagnetic potential (Lorenz gauge)	$\square^2\mathbf{A}=\mu_0\mathbf{J}$ $\square^2\varphi=\rho/\varepsilon_0$	$\partial_\mu\partial^\mu A^\nu=\mu_0 J^\nu$	$\square^2 A=\mu_0 J$
Electromagnetic fields (any gauge)	$\mathbf{E}=-\boldsymbol{\nabla}\varphi-\frac{\partial\mathbf{A}}{\partial t}$ $\mathbf{B}=\boldsymbol{\nabla}\times\mathbf{A}$	$F^{\mu\nu}=2\partial^{[\mu}A^{\nu]}$	$F=\square\wedge A$
Maxwell's equations	$\boldsymbol{\nabla}\cdot\mathbf{D}=\rho$ $\boldsymbol{\nabla}\times\mathbf{H}-\frac{\partial\mathbf{D}}{\partial t}=\mathbf{J}$ $\boldsymbol{\nabla}\cdot\mathbf{B}=0$ $\boldsymbol{\nabla}\times\mathbf{E}+\frac{\partial\mathbf{B}}{\partial t}=0$	$\partial_\mu F^{\mu\nu}=\mu_0 J^\mu$ $\partial_{[\mu}F_{\nu\sigma]}=0$, or $\partial_\nu G^{\mu\nu}=0$	$\square F=\mu_0 J$
Continuity equation	$\boldsymbol{\nabla}\cdot\mathbf{J}+\frac{\partial\rho}{\partial t}=0$	$\partial_\mu J^\mu=0$	$\square\cdot J=0$

This is the inhomogeneous Maxwell's equation. The homogeneous equation is found by considering the spacetime curl,

$$\square\wedge F=\square\wedge(\square\wedge A)=(\square\wedge\square)\wedge A=0 \tag{5.83}$$

where we have exploited the associativity of the wedge product, and the final operator in parentheses (like any wedge product of identical vectors) is zero. Thus, by adding (5.83) to (5.82), we obtain

$$\square\cdot F+\square\wedge F=\mu_0 J+0 \tag{5.84a}$$

$$\therefore\ \square F=\mu_0 J \tag{5.84b}$$

That is all! The use of the singular "Maxwell's equation" in the heading of this section was no accident; all of the fundamental laws of electromagnetics boil down to this one, exceedingly simple equation [10].

These results are summarized in Table 5.2.

5.2.5 Plane Waves

Electrical engineers are trained to make extensive use of spectral analysis, leveraging complex numbers to simplify the representation of time-harmonic forms. Historically, there was nothing fundamental in the laws of electromagnetics that demanded the use of complex numbers, they were simply convenient in that they replicated the behavior, through Euler's formula, of something that we wished to describe, namely sinusoidal waveforms. As we shall see, spacetime algebra makes such ad hoc introductions of complex numbers unnecessary, since it furnishes us automatically with an imaginary unit having a concrete physical interpretation as a four-volume in spacetime, and whose appearance in equations involving sinusoidal waves, it turns out, is inevitable [1].

Example: In vacuum, the canonical polar form for bivectors not only serves as the frame-independent decomposition of the Faraday field, but becomes the ansatz for the solution to Maxwell's equation itself,

$$F = \mathbf{f}e^{i\varphi} \tag{5.85}$$

With the goal of generating plane wave solutions (wherein the wavefront is a plane extending infinitely in all directions perpendicular to the line of propagation), we will assume that the simple bivector coefficient, $\mathbf{f}$, is constant for all spacetime coordinates, so that all variations are contained in the phase factor, φ (sometimes called the complexion [11] of F)

$$\Box F = \Box\left(\mathbf{f}e^{i\varphi}\right) = (\Box\varphi)i\mathbf{f}e^{i\varphi} = 0 \tag{5.86a}$$

$$\therefore (\Box\varphi)\mathbf{f} = 0 \tag{5.86b}$$

We cannot have $\mathbf{f} = 0$, for then the entire field would be zero everywhere, which is trivial. Nor can we allow $\Box\varphi = 0$, for then the field would be unchanging throughout spacetime, a static field that does not correspond to the wave solution we seek. From this, we conclude that both $\mathbf{f}$ and $\Box\varphi$ must be null (nonzero in value but having zero square norm), for otherwise we could invert one in (5.86b) and determine that the other is zero-valued.

Since $\mathbf{f}$ is simple, we can write it as the wedge product of two vectors,

$$\mathbf{f} = s \wedge L \tag{5.87}$$

Furthermore, because $\mathbf{f}$ is null, one of these two vectors must also be null. Let that vector be L. The other vector, s, is selected to be orthogonal to the first, $s \cdot L = 0$, so that

$$sL = s \cdot L + s \wedge L = s \wedge L = \mathbf{f} \tag{5.88}$$

Turning our attention now to $\Box\varphi$, we conclude that since $\mathbf{f}$ is constant then $\Box\varphi$ must also be constant, for else (5.86b) could not be satisfied for all x. This suggests

$$\varphi = K \cdot x + \varphi_0 \tag{5.89}$$

where $K = \Box\varphi$ is a constant null vector, and the dot product is chosen since φ must be scalar. Substituting this into (5.86b), we have

$$(\Box\varphi)\mathbf{f} = K\mathbf{f} = KsL = 0 \tag{5.90}$$

which is satisfied if $L = K$ (we cannot have $s = K$, because L and K are required to be null while s is not).

We thus have all the terms of a plane wave solution in spacetime,

$$F = sKe^{i(K \cdot x + \varphi_0)} \tag{5.91}$$

The vector K is of course the four-wavevector and $K \cdot x$ is the Lorentz-invariant phase that we have made reference to elsewhere in this book. That K is null proves the wave is lightlike, traveling at speed c.

What is new and perhaps surprising here is that there is no need to throw away the imaginary part of this equation, as is traditionally done with the phasor representation of wave solutions; both the real and imaginary parts of this equation have physical meaning. To see what they are, let us expand this equation in the reference frame of a particular observer selected for simplicity: s will be directed toward $\mathbf{x}$, propagation will be in the $\mathbf{z}$ direction, and the time reference will be selected to eliminate the phase offset,

$$s = s^1 \gamma_1 \tag{5.92a}$$

$$K\gamma_0 = \tfrac{\omega}{c} + \mathbf{k} = \tfrac{\omega}{c} + k\sigma_3 = k\,(1 + \sigma_3) \tag{5.92b}$$

$$\varphi_0 = 0 \tag{5.92c}$$

Therefore,

$$F = sKe^{iK \cdot x} = \left(s^1\gamma_1\right) k\,(\gamma_0 + \gamma_3)\, e^{i(\omega t - kz)} \tag{5.93a}$$

$$= s^1 k\,(\sigma_1 + i\sigma_2)\, e^{i(\omega t - kz)} = s^1 k\,(\mathbf{x} + i\mathbf{y})\, e^{i(\omega t - kz)} = \tfrac{1}{c}\mathbf{E} + i\mathbf{B} \tag{5.93b}$$

where

$$\mathbf{E} = cs^1 k \left[\mathbf{x}\cos(\omega t - kz) - \mathbf{y}\sin(\omega t - kz)\right] \tag{5.94a}$$

$$\mathbf{B} = s^1 k \left[\mathbf{x}\sin(\omega t - kz) + \mathbf{y}\cos(\omega t - kz)\right] \tag{5.94b}$$

These are formulas for a circularly polarized plane wave. The real part of the bivector yielded the polar vector $\mathbf{E}$, while the imaginary part yielded the axial vector $\mathbf{B}$. This is in contrast to the traditional method, wherein separate complex phasor equations are written for both $\mathbf{E}$ and $\mathbf{B}$, they are constrained to be proportional to one another in accordance with the wave impedance of free space, and then the imaginary part of each equation is discarded.

It is interesting to note that, while the ansatz we chose in (5.85) incorporated the grade-4 pseudoscalar within the exponential phase argument, another selection, a bivector, is possible that leads to the same result [12]. We could instead have used

$$F = \mathbf{f}e^{\gamma_{12}\varphi} = s^1 k \left(\sigma_1 + i\sigma_2\right)\left(\cos\varphi + \gamma_{12}\sin\varphi\right) \tag{5.95a}$$

$$= s^1 k \left(\mathbf{x}\cos\varphi - \mathbf{y}\sin\varphi\right) + is^1 k \left(\mathbf{x}\sin\varphi + \mathbf{y}\sin\varphi\right) \tag{5.95b}$$

which is the same as (5.94) when we expand $\varphi = K\cdot x = \omega t - kz$. The criterion for a template $\mathbf{f}e^{q\varphi}$ where $q^2 = -1$ to match our solution above is that $\mathbf{f}q = \mathbf{f}i$. Since $\mathbf{f}$ is null, it cannot be divided out to guarantee that $q = i$, and thus both $q = i$ and $q = \gamma_{12}$ are possible solutions. I prefer the pseudoscalar i since its commutation properties are simpler (i commutes with all bivectors; γ_{12} commutes with some bivectors but not others). In addition, $i\varphi$ is invariant under Lorentz transformation, while $\gamma_{12}\varphi$ is not.

Circular polarization of the opposite handedness can be selected by conjugating the argument of the exponential,

$$F = sKe^{-iK\cdot x} \tag{5.96}$$

while the direction of propagation can be reversed by negating the spatial three-vector $\mathbf{k}$ (not the full four-vector K). Linear polarization, then, is given by the superposition of two spacetime waves having opposite circular polarization,

$$F_{lin} = \tfrac{1}{2}sK\left(e^{iK\cdot x} + e^{-iK\cdot x}\right) = sK\cos(K\cdot x) \tag{5.97}$$

Note that F is still complex (because sK is complex) representing the polar and axial components of the Faraday field, only now they are in-phase in all reference frames.

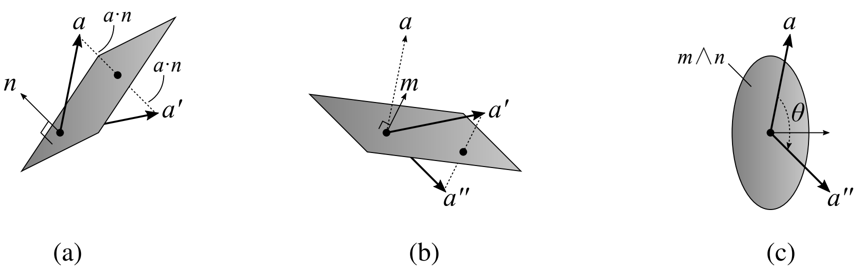


Figure 5.7 Construction of a rotation by two reflections. (a) First reflection of vector a through hyperplane normal to n. (b) Second reflection through hyperplane normal to m. (c) Equivalent rotation defined by the magnitude and orientation of the bivector $m \wedge n$.

5.3 TRANSFORMATIONS

What of Lorentz transformations, for example, rotations and boosts? How are we to write these operations, which are so central to electromagnetics in the context of special relativity, using the formalism of spacetime algebra? Each of these, we shall see, takes on the simple form of a rotation.

5.3.1 Reflection

First, let us consider a simple reflection in the hyperplane normal to the unit four-vector, n, as shown in Figure 5.7(a) (remember that this is a three-dimensional visualization of a process that occurs in the four dimensions of spacetime). The reflection of the vector a through the hyperplane may be found by twice subtracting the projection of a onto the unit vector n,

$$a' = a - 2(a \cdot n)n = a - (an + na)n = a - an^2 - nan = -nan \tag{5.98}$$

where we have used the geometric product to simplify the result.

5.3.2 Rotation

As noted in Section 2.1.3, reflection is an improper physical transformation, inverting geometric features and transforming axial vectors in a manner that is not consistent with physical laws. A proper rotation can instead be formed by two reflections back-to-back, restoring the handedness of the coordinate system. We thus apply a second reflection using the same formula derived above, but with a new hyperplane perpendicular to the unit vector m, as illustrated in Figure 5.7(b),

$$a'' = -ma'm = -m(-nan)m = (mn)a(mn)^{\sim} \tag{5.99}$$

where the superscript tilde ($^{\sim}$) indicates reversion, or the reverse-ordering of vectors in a geometric product. Applied to general multivectors, reversion negates the sign of their component bivectors and trivectors (i.e., pseudovectors), while leaving scalars, vectors, and pseudoscalars unchanged [1, 2].

The multivector (or mixed-grade) quantity $R = mn$ is called a rotor, characterized by the normalization,

$$R\widetilde{R} = mnnm = 1 \tag{5.100}$$

(Similar forms which do not obey this normalization are more generally called spinors.) Rotors can be written as the exponential of a bivector, G,

$$R = e^{+G/2} \tag{5.101a}$$

$$\widetilde{R} = e^{-G/2} \tag{5.101b}$$

where G is called the generator of the rotation, and describes the spacetime plane in which rotation occurs.

Take, for example, a rotor formed from the generator $G = \theta\gamma_{21}$ applied to the vector a,

$$a'' = Ra\widetilde{R} = e^{G/2}ae^{-G/2} = e^{\theta\gamma_{21}/2}ae^{-\theta\gamma_{21}/2} \tag{5.102a}$$

$$= \left(\cos\tfrac{\theta}{2} + \gamma_{21}\sin\tfrac{\theta}{2}\right) a \left(\cos\tfrac{\theta}{2} - \gamma_{21}\sin\tfrac{\theta}{2}\right) \tag{5.102b}$$

$$= a\cos^2\tfrac{\theta}{2} + (\gamma_{21}a - a\gamma_{21})\sin\tfrac{\theta}{2}\cos\tfrac{\theta}{2} - \gamma_{21}a\gamma_{21}\sin^2\tfrac{\theta}{2} \tag{5.102c}$$

$$= \tfrac{1}{2}(a - \gamma_{21}a\gamma_{21}) + \tfrac{1}{2}(a + \gamma_{21}a\gamma_{21})\cos\theta + \tfrac{1}{2}(\gamma_{21}a - a\gamma_{21})\sin\theta \tag{5.102d}$$

where we have used the Euler expansion (5.28) since $\gamma_{21}^2 = -1$. We may further note that γ_{21} commutes with γ_0 and γ_3, but anticommutes with γ_1 and γ_2. Therefore, expanding a in terms of its basis vectors, $a^k\gamma_k$,

$$a'' = a^0\gamma_0 + \left(a^1\gamma_1 + a^2\gamma_2\right)\cos\theta - \left(a^1\gamma_1 + a^2\gamma_2\right)\gamma_{21}\sin\theta + a^3\gamma_3 \tag{5.103a}$$

$$= a^0\gamma_0 + \left(a^1\gamma_1 + a^2\gamma_2\right)\cos\theta - \left(a^1\gamma_2 - a^2\gamma_1\right)\sin\theta + a^3\gamma_3 \tag{5.103b}$$

$$= a^0\gamma_0 + \left(a^1\cos\theta + a^2\sin\theta\right)\gamma_1 + \left(a^2\cos\theta - a^1\sin\theta\right)\gamma_2 + a^3\gamma_3 \tag{5.103c}$$

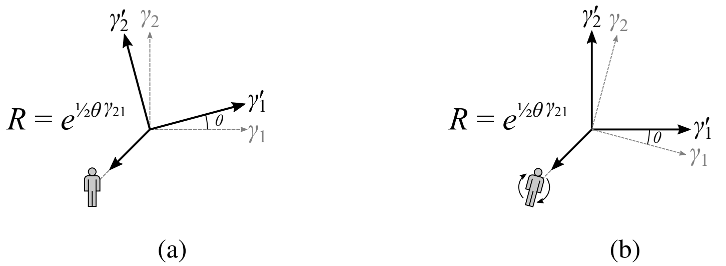


Figure 5.8 The effect of a rotation using rotor $R = e^{\theta\gamma_{21}/2}$. (a) In the observer's reference frame, objects seem to rotate from γ_1 toward γ_2. (b) In the global reference frame, the observer himself rotates from γ_2 toward γ_1.

or, in matrix form,

$$\mathbf{a}'' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & \sin\theta & 0 \\ 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a^0 \\ a^1 \\ a^2 \\ a^3 \end{pmatrix} \tag{5.104}$$

This has clearly revealed a rotation of the coordinate axes through an angle θ from the γ_1 axis toward the γ_2 axis in the γ_{21} plane, as illustrated in Figure 5.8(a). Alternatively, it can be visualized as the rotation of the observer's reference frame in the opposite direction, from γ_2 toward γ_1, as shown in Figure 5.8(b).

5.3.3 Lorentz Boosts

We purposefully selected a generator spanning two spacelike axes (γ_1 and γ_2) in the previous example. Suppose we had chosen the generator $G = \zeta\gamma_{01}$, involving the timelike axis, γ_0, instead? In that case,

$$a'' = Ra\widetilde{R} = e^{G/2} a e^{-G/2} = e^{\zeta\gamma_{01}/2} a e^{-\zeta\gamma_{01}/2} \tag{5.105a}$$

$$= \left(\cosh\tfrac{\zeta}{2} + \gamma_{01}\sinh\tfrac{\zeta}{2}\right) a \left(\cosh\tfrac{\zeta}{2} - \gamma_{01}\sinh\tfrac{\zeta}{2}\right) \tag{5.105b}$$

$$= a\cosh^2\tfrac{\zeta}{2} + (\gamma_{01}a - a\gamma_{01})\sinh\tfrac{\zeta}{2}\cosh\tfrac{\zeta}{2} - \gamma_{01}a\gamma_{01}\sinh^2\tfrac{\zeta}{2} \tag{5.105c}$$

$$= \tfrac{1}{2}(a + \gamma_{01}a\gamma_{01}) + \tfrac{1}{2}(a - \gamma_{01}a\gamma_{01})\cosh\zeta + \tfrac{1}{2}(\gamma_{01}a - a\gamma_{01})\sinh\zeta \tag{5.105d}$$

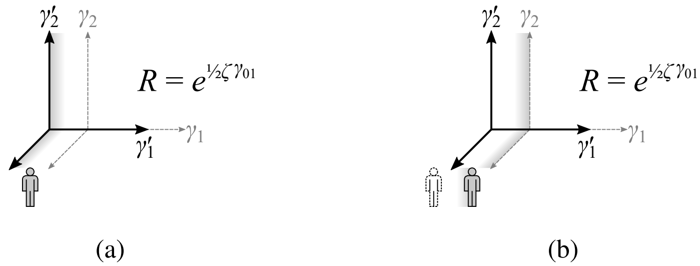


Figure 5.9 The effect of a boost using rotor $R = e^{\zeta\gamma_{01}}$. (a) In the observer's reference frame, objects seem to move in the $-\gamma_1$ direction. (b) In the global reference frame, the observer moves in the $+\gamma_1$ direction.

where we have used the hyperbolic functions since $\gamma_{01}^2 = 1$. Once again expanding $a = a^k\gamma_k$, and noting that γ_{01} anticommutes with γ_0 and γ_1, but commutes with γ_2 and γ_3,

$$a'' = \left(a^0\gamma_0 + a^1\gamma_1\right)\cosh\zeta - \left(a^0\gamma_0 + a^1\gamma_1\right)\gamma_{10}\sinh\zeta + a^2\gamma_2 + a^3\gamma_3 \quad (5.106a)$$

$$= \left(a^0\gamma_0 + a^1\gamma_1\right)\cosh\zeta - \left(a^0\gamma_1 + a^1\gamma_0\right)\sinh\zeta + a^2\gamma_2 + a^3\gamma_3 \quad (5.106b)$$

$$= \left(a^0\cosh\zeta - a^1\sinh\zeta\right)\gamma_0 + \left(a^1\cosh\zeta - a^0\sinh\zeta\right)\gamma_1 + a^2\gamma_2 + a^3\gamma_3 \quad (5.106c)$$

or, equivalently,

$$\mathbf{a}'' = \begin{pmatrix} \cosh\zeta & -\sinh\zeta & 0 & 0 \\ -\sinh\zeta & \cosh\zeta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a^0 \\ a^1 \\ a^2 \\ a^3 \end{pmatrix} \quad (5.107)$$

We see that this operation has affected a Lorentz boost of the observational reference frame in the γ_1 direction with rapidity ζ. To the observer in this moving reference frame, objects appear to move in the $-\gamma_1$ direction. These effects are illustrated in Figure 5.9.

Transformation of the Faraday field can be derived by applying the above rotors to the spacetime gradient and four-potential (both vectors) on which the field is defined,

$$F' = \square' \wedge A' = \left(R\square\widetilde{R}\right) \wedge \left(RA\widetilde{R}\right) \quad (5.108a)$$

$$= \tfrac{1}{2}\left(R\square\widetilde{R}RA\widetilde{R} - R\dot{A}\widetilde{R}R\dot{\square}\widetilde{R}\right) = \tfrac{1}{2}\left(R\square A\widetilde{R} - R\dot{A}\dot{\square}\widetilde{R}\right) \quad (5.108b)$$

$$= \tfrac{1}{2}R\left(\Box A - \dot{A}\dot{\Box}\right)\widetilde{R} = R(\Box \wedge A)\widetilde{R} = RF\widetilde{R} \tag{5.108c}$$

Thus, we see that a grade-2 bivector in spacetime algebra transforms in the same way as a grade-1 vector. In fact, any multivector, $\mathcal{M}$, transforms in exactly the same way, $\mathcal{M}' = R\mathcal{M}\widetilde{R}$.

A corollary to this is that proper scalars, or any quantities with which R commutes, are invariant under transformation. Consider the wave phase term, $\Phi = K \cdot x$, for example. Elsewhere, we have referred to this as the *Lorentz-invariant* phase, an apt description since

$$K' \cdot x' = \left(RK\widetilde{R}\right) \cdot \left(Rx\widetilde{R}\right) = R(K \cdot x)\widetilde{R} = R\widetilde{R}(K \cdot x) = K \cdot x \tag{5.109}$$

where the second-to-last step follows since $K \cdot x$ is a scalar. More simply,

$$\Phi' = R\Phi\widetilde{R} = R\widetilde{R}\Phi = \Phi \tag{5.110}$$

Pseudoscalars also inevitably commute with all rotors, R, and are likewise invariant under Lorentz transformation. Scalars and pseudoscalars may sometimes be referred to as the two *invariant grades* for this reason.

It is instructive to decompose the Faraday field into parts that are parallel and perpendicular to the direction of the boost,

$$F = F_{\|} + F_{\perp} \tag{5.111a}$$

$$\therefore F' = RF\widetilde{R} = e^{\zeta G/2}\left(F_{\|} + F_{\perp}\right)e^{-\zeta G/2} \tag{5.111b}$$

We can simplify this by noting that G commutes with the parallel field components, while anticommuting with the perpendicular components. For example, $G = \gamma_{01} = -\sigma_1$ commutes with σ_1 and $i\sigma_1$, but anticommutes with σ_2, $i\sigma_3$, and so on. This gives us the following rotor commutation rules

$$F_{\|}e^{-\zeta G/2} = e^{-\zeta G/2}F_{\|} \tag{5.112a}$$

$$F_{\perp}e^{-\zeta G/2} = e^{+\zeta G/2}F_{\perp} \tag{5.112b}$$

Therefore,

$$F' = F_{\|} + e^{\zeta G}F_{\perp} = F_{\|} + (\cosh\zeta + G\sinh\zeta)F_{\perp} \tag{5.113}$$

This derivation makes it clear that only the transverse components of the Faraday field (whether the associated spatial vectors are polar or axial) are affected by a boost.

A summary of the two general classes of rotors and the kind of transformations they bring about is given in Table 5.3.

Table 5.3
Reference Frame Transformations in Spacetime Algebra
Form: $\mathcal{M}' = R\mathcal{M}\widetilde{R}$, where $R = e^{G/2}$

Type	Generator, G	Direction	Amount
Rotation	$\theta\gamma_{kl}$	from γ_k to γ_l	angle θ
Boost	$\zeta\gamma_{0k}$	toward γ_k	rapidity ζ

Note: γ_0 is the timelike axis, while γ_k and γ_l are spacelike axes.

5.3.4 Compound Transformations

We have thus shown that all permissible Lorentz transformations take the form of rotors in spacetime algebra, exponentiating bivector generators corresponding to the planes of rotation — perpendicular to the timelike axis, γ_0, for purely spatial rotations, and parallel to it for boosts. When two rotors, R_1 and R_2, are combined, we have

$$a' = R_2 R_1 a \widetilde{R}_1 \widetilde{R}_2 = R_2 R_1 a \left(R_2 R_1\right)^{\sim} = R a \widetilde{R} \tag{5.114}$$

where

$$R = R_2 R_1 \tag{5.115}$$

We might be tempted to further write that

$$R_2 R_1 = e^{G_2} e^{G_1} \stackrel{?}{=} e^{G_2 + G_1} = e^{G_1 + G_2} \stackrel{?}{=} e^{G_1} e^{G_2} = R_1 R_2 \tag{5.116}$$

but this is incorrect. The product-of-powers rule (the equalities tagged with question marks) is only valid when the exponentiated quantities commute. More generally, evaluation of the product of powers must take into account the *Baker-Campbell-Hausdorff formula* [13],

$$e^X e^Y = e^{X+Y+[X,Y]+\frac{1}{3}([X,[X,Y]]+[Y,[Y,X]])+\cdots} \tag{5.117}$$

where $[X, Y] = \frac{1}{2}(XY - YX)$ is called the *commutator bracket* or *commutator product* of X and Y (see Appendix D). When the two exponentiated terms commute ($XY = YX$), this complex expression reduces to the simpler form we are most familiar with from scalar mathematics. In the more general case, G_1 and G_2 do not commute, and the outcome of multiple rotations or boosts is typically dependent upon the sequence.

5.4 SUBALGEBRAS

It is interesting to note that any operation defined in spacetime algebra acting upon scalars, pseudoscalars, and bivectors can only produce scalars, pseudoscalars, and bivectors as a result. These are merely the even-numbered grades, with basis components given by

$$\{1\}, \{\sigma_k\}, \{i\sigma_k\}, \{i\} \tag{5.118}$$

forming what is referred to as the *even subalgebra* of the original, complete set including vectors and pseudovectors. We have already associated with σ_k the spatial three-vectors, and the operations that we normally apply to such objects have already been defined, especially the dot product $(\mathbf{a}\cdot\mathbf{b})$ and the cross product $(\mathbf{a}\times\mathbf{b})$, the latter indirectly as the Hodge dual of the commutator bracket,

$$\mathbf{a}\times\mathbf{b} = [\mathbf{a},\mathbf{b}]\, i^{-1} = -\tfrac{i}{2}(\mathbf{ab}-\mathbf{ba}) \tag{5.119}$$

Thus, all of the formalism of classical, three-dimensional vector calculus, with which the reader is likely most familiar, complete with the richness of complex analysis, is included as a subset of spacetime algebra. The spacetime split effectively maps frame-independent vector quantities in spacetime to the even subalgebra experienced by an observer in the reference frame defined by the worldine γ_0 [6]. Now, however, the cross product that we use every day in our profession (at least implicitly) is defined in a way that is consistent and extensible to higher dimensions, while the imaginary unit, i, has been supplied with a geometric interpretation as the unique, unit four-volume in Minkowski spacetime.

Mathematicians denote the spacetime algebra with the metric $(+---)$ as $\mathcal{C}_{1,3}(\mathbb{R})$, where $\mathcal{C}_{1,3}$ recognizes it as a Clifford algebra (that is, an algebra equipped with a Clifford product) built upon a fundamental vector space having one basis vector with positive square norm (γ_0) and three with negative square norm ($\gamma_{1,2,3}$), and $\mathbb{R}$ specifies that the coefficients of those vectors must be real numbers. The scalars, bivectors, trivectors, and pseudoscalars which also comprise the domain of spacetime algebra are derived elements formed by the Clifford products of those four fundamental vectors. The even subalgebra, then, would be annotated $\mathcal{C}_{3,0}(\mathbb{R})$, as it has only three fundamental vectors, σ_k, all having positive square norm, and the associated scalar, pseudoscalar, and pseudovector products which can be derived from them,

$$1 = \sigma_k^2 \tag{5.120a}$$

$$i = \sigma_1\sigma_2\sigma_3 \tag{5.120b}$$

$$i\sigma_k = \sigma_1\sigma_2\sigma_3\sigma_k \tag{5.120c}$$

Continuing in this fashion, the even subalgebra of three-dimensional vector calculus is written $\mathcal{C}_{0,2}(\mathbb{R})$ and may be associated with Hamilton's (now largely abandoned) quaternion mathematics [14]. The even subalgebra of quaternions is $\mathcal{C}_{0,1}(\mathbb{R})$ representing scalar complex analysis, and the even subalgebra of that is $\mathcal{C}_{0,0}(\mathbb{R})$ which is real number theory. Working backwards, it can be shown that spacetime algebra itself is the subalgebra of $\mathcal{C}_{4,1}(\mathbb{R})$ and that of $\mathcal{C}_{2,4}(\mathbb{R})$, which correspond to relativistic electron physics and conformal space, respectively [1].

We thus see that spacetime algebra (underlined in the sequence below) finds its natural place in a nested embedding of various formalisms within mathematical physics which were discovered (or engineered) at different times in history to describe aspects of the rich and elegant universe in which we find ourselves,

$$\mathcal{C}_{2,4}(\mathbb{R}) \supset \mathcal{C}_{4,1}(\mathbb{R}) \supset \underline{\mathcal{C}_{1,3}(\mathbb{R})} \supset \mathcal{C}_{3,0}(\mathbb{R}) \supset \mathcal{C}_{0,2}(\mathbb{R}) \supset \mathcal{C}_{0,1}(\mathbb{R}) \supset \mathcal{C}_{0,0}(\mathbb{R}) \tag{5.121}$$

Indeed, the choice of many of the symbols we have used are borrowed from these historical approaches:

- The unit four-volume i references its algebraic identification with complex analysis, $\mathcal{C}_{0,1}(\mathbb{R})$.
- γ_μ is taken from Dirac algebra [15] for spin-$\frac{1}{2}$ particles, conforming algebraically to $\mathcal{C}_{1,3}(\mathbb{R})$.
- σ_k derives from the three Pauli matrices [16] used to describe particle spin in the presence of an electromagnetic field. These are now understood to be basis components of $\mathcal{C}_{3,0}(\mathbb{R})$.

That $\mathcal{C}_{1,3}(\mathbb{R})$ rather than $\mathcal{C}_{3,1}(\mathbb{R})$ arises in this chain of embedded algebraic systems is at least a philosophical justification for choosing the metric that we have — (+ – – –) instead of (– + + +) — despite the lack of measurable physical significance of either.

References

[1] J. Dressel, K. Y. Bliokh, and F. Nori, "Spacetime algebra as a powerful tool for electromagnetism," *Physics Reports*, vol. 589, pp. 1–71, 2015.

[2] C. Doran, A. Lasenby, S. Gull, S. Somaroo, and A. Challinor, "Spacetime algebra and electron physics," *Advances in Imaging and Electron Physics*, vol. 95, pp. 271–386, 1996.

[3] Wikipedia. (2024, January) Spacetime algebra. https://en.wikipedia.org/wiki/Spacetime_algebra.

[4] J. Lasenby, A. N. Lasenby, and C. J. L. Doran, “A unified mathematical language for physics and engineering in the 21st century,” *Philosophical Transactions of the Royal Society of London A*, vol. 358, no. 1765, pp. 21–39, January 2000.

[5] D. Hestenes, “A unified language for mathematics and physics,” *Clifford Algebras and Their Applications in Mathematical Physics*, pp. 1–23, 1986.

[6] D. Hestenes. (1998) Spacetime calculus. http://geocalc.clas.asu.edu/pdf/SpaceTimeCalc.pdf.

[7] A. Arsenovic, “Applications of conformal geometric algebra to transmission line theory,” *IEEE Access*, vol. 5, pp. 19 920–19 941, July 2017.

[8] Wikipedia. (2023, November) Bivector. https://en.wikipedia.org/wiki/Bivector.

[9] J. Soni and R. Goodman, *A Mind at Play: How Claude Shannon Invented the Information Age*. New York: Simon & Schuster, 2017.

[10] Wikipedia. (2024, January) Geometric algebra. https://en.wikipedia.org/wiki/Geometric_algebra.

[11] D. Hestenes, *Space-Time Algebra*, 2nd ed. Switzerland: Birkhäuser, 2015.

[12] A. Arsenovic, “Spacetime transmission lines,” *in preparation*, 2020.

[13] Wikipedia. (2023, November) Baker-Campbell-Hausdorff formula. https://en.wikipedia.org/wiki/Baker%E2%80%93Campbell%E2%80%93Hausdorff_formula.

[14] W. R. Hamilton, *Lectures on Quaternions*. Dublin: Hodges and Smith, 1853.

[15] Wikipedia. (2023, September) Dirac algebra. https://en.wikipedia.org/wiki/Dirac_algebra.

[16] Wikipedia. (2023, December) Pauli matrices. https://en.wikipedia.org/wiki/Pauli_matrices.

Chapter 6

Interactions with Matter

The preceding treatment of electrodynamics in the context of special relativity has been limited to fields existing in a vacuum, for we previously argued that even in the presence of dielectric media, the fields themselves would always exist in the microscopic gaps between free and bound molecular charges. This allowed us to derive some fundamental, if unexpected, properties of space and time. However, in order to harness these principles for engineering, we must be able to control the Faraday tensor or bivector field by manipulating boundary conditions. At some point, we have to account for the interaction of the field with solid matter.

6.1 MACROSCOPIC FIELD EQUATIONS

In Chapter 1, we identified the vacuum permittivity, ε_0, and vacuum permeability, μ_0, as the constituent parameters relating the electric field to the electric displacement and the magnetic field to the magnetic flux density, respectively,

$$\mathbf{D} = \varepsilon_0 \mathbf{E} \tag{6.1a}$$

$$\mathbf{B} = \mu_0 \mathbf{H} \tag{6.1b}$$

As long as the fields existed in vacuum, we said, these relationships have the character of unit conversions, for there is no medium to account for the scaling; they are simply a matter of definition.

These definitions hold true in the presence of matter as well, so long as we take a microscopic view, wherein the bound charges and magnetic dipole moments of the molecular lattice are seen as independent sources making their own contributions to the electromagnetic fields. However, it is frequently more convenient in engineering work to take a macroscopic view, wherein the material is seen as continuous, electrically neutral, and inert insofar as the sources of electromagnetic fields is concerned [1–3].

6.1.1 Material Response Bivectors

We saw in Section 1.4.1 that the impact of bound charges and currents within a material can be subsumed into the constituent parameters by modifying their values,[1]

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} = \varepsilon_0 \left(1 + \chi_e\right) \mathbf{E} = \varepsilon_0 \varepsilon_r \mathbf{E} = \varepsilon \mathbf{E} \tag{6.2a}$$

$$\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M}) = \mu_0 \left(1 + \chi_m\right) \mathbf{H} = \mu_0 \mu_r \mathbf{H} = \mu \mathbf{H} \tag{6.2b}$$

In these equations, the dielectric polarization density, $\mathbf{P}$, and the magnetization, $\mathbf{M}$, represent the electric displacement and magnetic field associated with the bound charges and currents inside the material, whereas the traditional displacement and magnetic fields account only for free charges and currents. The complete field bivector, then, is simply the superposition of the two.

To make this more explicit, let us separate the inhomogeneous Maxwell's equation into two parts, one associated with free sources, and the other with bound sources,

$$\Box \cdot F = \mu_0 J \tag{6.3a}$$

$$\Box \cdot (\mathcal{D} + \mathcal{M}) = J_{free} + J_{bound} \tag{6.3b}$$

where the first term inside the parentheses, $\mathcal{D}$, is the *electromagnetic displacement bivector*, and is associated with free sources only, while the second term, $\mathcal{M}$, is the *magnetization-polarization bivector*, and is associated with bound sources only [4]. Therefore, we may write

$$\Box \cdot \mathcal{D} = J_{free} \tag{6.4a}$$

$$\Box \cdot \mathcal{M} = J_{bound} \tag{6.4b}$$

where, in the canonical form,

$$\mathcal{D} = \mathbf{d} e^{i\psi} \tag{6.5a}$$

$$\mathcal{M} = \mathbf{m} e^{i\xi} \tag{6.5b}$$

---

1 We will assume for convenience here that the material is linear, isotropic, and nondispersive. The constituent parameters, ε and μ, are therefore scalar quantities.

or, in the rest frame of the material,

$$\mathcal{D} = c\mathbf{D} + i\mathbf{H} \tag{6.6a}$$

$$\mathcal{M} = -c\mathbf{P} + i\mathbf{M} \tag{6.6b}$$

and

$$\mu_0(\mathcal{D} + \mathcal{M}) = \mu_0 c(\mathbf{D} - \mathbf{P}) + i\mu_0(\mathbf{H} + \mathbf{M}) = \tfrac{1}{c}\mathbf{E} + i\mathbf{B} = F \tag{6.7}$$

matching the definitions in (6.2).

6.1.2 Conservation of Free and Bound Charge

The form of (6.4) guarantees that free and bound charge are both conserved independently; bound charge cannot be freed and free charge cannot be bound simply by application of external fields (at least, not within the material's linear regime). To see this, we need only take the spacetime divergence of each,

$$\Box \cdot J_{free} = \Box \cdot (\Box \cdot \mathcal{D}) = 0 \tag{6.8a}$$

$$\Box \cdot J_{bound} = \Box \cdot (\Box \cdot \mathcal{M}) = 0 \tag{6.8b}$$

In both cases, we have applied the identity (D.168).

We shall have little more to say about bound charges and currents throughout the rest of this book, and will for convenience omit the label "free" when writing the four-current density, except when required for clarity. It should be understood from context that the displacement field and associated macroscopic forms couple only to the free sources within a material.

6.1.3 Constitutive Relations

Let us attempt to write a constitutive relationship between the displacement field, $\mathcal{D}$, and the Faraday field, F, similar to the classical constitutive relationships in (6.2). First, we note that the electric and magnetic fields may be isolated by taking the dot and wedge products of F with a timelike axis, γ_0,

$$F \cdot \gamma_0 = \left(\tfrac{1}{c}\mathbf{E} + i\mathbf{B}\right) \cdot \gamma_0 = \tfrac{1}{c}\mathbf{E}\gamma_0 \tag{6.9a}$$

$$F \wedge \gamma_0 = \left(\tfrac{1}{c}\mathbf{E} + i\mathbf{B}\right) \wedge \gamma_0 = i\mathbf{B}\gamma_0 \tag{6.9b}$$

(These results are easily proven by expanding the dot and wedge products via (5.34a) and (5.34b), and then considering the commutivity of the elements.) The explicit appearance of a basis vector (such as γ_0) in any spacetime-algebraic equation is a sign that we have tied our derivation to a particular reference frame. The macroscopic constituent parameters, μ and ε, that we apply next are valid only in the rest frame of the material, and are known as the *proper permeability* and *proper permittivity*, respectively[2] [4]. With that in mind, the displacement field in the rest frame may be written

$$\mathcal{D} = c\mathbf{D} + i\mathbf{H} = c\varepsilon\mathbf{E} + \tfrac{i}{\mu}\mathbf{B} = c^2\varepsilon\left(F \cdot \gamma_0\right)\gamma_0 + \tfrac{1}{\mu}\left(F \wedge \gamma_0\right)\gamma_0 \tag{6.10a}$$

$$= c^2\varepsilon\left(F \cdot (c\gamma_0)\right)\tfrac{\gamma_0}{c} + \tfrac{1}{\mu}\left(F \wedge (c\gamma_0)\right)\tfrac{\gamma_0}{c} \tag{6.10b}$$

where we have assumed a linear, isotropic material (in which μ and ε are scalars). Once again, the explicit presence of the basis vector, γ_0, restricts the use of this equation to a particular reference frame, specifically the rest frame of the material in which the macroscopic permittivity, ε, and permeability, μ, are defined. That restriction can be circumvented by recognizing that $c\gamma_0$ is also the four-velocity of the material in its own rest frame,

$$U = \left(c\gamma_0 + v^1\gamma_1 + v^2\gamma_2 + v^3\gamma_3\right)\cosh\zeta = c\gamma_0 \tag{6.11}$$

because, in the rest frame, $\mathbf{v} = 0$ and $\cosh\zeta = 1$. Furthermore,

$$\frac{\gamma_0}{c} = \frac{c\gamma_0}{c^2} = \frac{U}{U^2} = U^{-1} \tag{6.12}$$

Thus, we may write

$$\mathcal{D} = c^2\varepsilon(F \cdot U)U^{-1} + \tfrac{1}{\mu}(F \wedge U)U^{-1} = c^2\varepsilon F_U + \tfrac{1}{\mu}F_{\not{U}} \tag{6.13}$$

One may recognize the projection and rejection operations in the formula above; the projection of F onto U is scaled by $c^2\varepsilon$ and the rejection of F from U is scaled by $1/\mu$. A (three-dimensional) representation of what it means to project a plane onto a line segment and to reject the plane from that line segment is given in Figure 6.1.

---

2 The material constituent parameters, permeability (μ) and permittivity (ε), will generally transform from one reference frame to the next. The term proper will be applied to these parameters and their dependents (e.g., the index of refraction (n) and wave impedance (η)) to indicate the values they take on in the rest frame of the material.

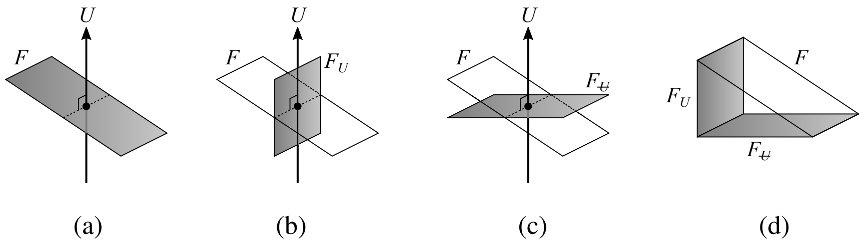


Figure 6.1 Illustration of the projection and rejection of a bivector plane, F, onto and from a vector, U. (a) Original vector and bivector. (b) The projection of F onto U. (c) The rejection of F from U. (d) Bivector addition, $F = F_U + F_{\not U}$.

Equation (6.13) represents the constitutive relation for the displacement field, and although its root derivation was performed in the rest frame of the material, its form guarantees that it transforms covariantly into any reference frame. Note that ε and μ directly relate $\mathbf{E}$ to $\mathbf{D}$ and $\mathbf{B}$ to $\mathbf{H}$ in the rest frame only. These field quantities become mixed in other reference frames, and can no longer be related by simple scalar constants — for one, it is well-known in classical electromagnetics that the parameters ε and μ determine the phase velocity of light in the medium, but in the relativistic analysis, the phase velocity in the medium must be direction-dependent due to the medium's motion relative to the reference frame of observation. This phenomenon will be explored in Section 6.2.

The constituent relationship can be inverted, thus solving for F in terms of $\mathcal{D}$, with the help of identities (D.151) and (D.152),

$$(\mathcal{D} \cdot U)U^{-1} = c^{-2}(\mathcal{D} \cdot U)U = c^{-2}(\mathcal{D}U) \wedge U = c^2\varepsilon(F \cdot U)U^{-1} \tag{6.14a}$$

$$(\mathcal{D} \wedge U)U^{-1} = c^{-2}(\mathcal{D} \wedge U)U = c^{-2}(\mathcal{D}U) \cdot U = \tfrac{1}{\mu}(F \wedge U)U^{-1} \tag{6.14b}$$

or, more compactly,

$$\mathcal{D}_U = c^2\varepsilon F_U \tag{6.15a}$$

$$\mathcal{D}_{\not U} = \tfrac{1}{\mu}F_{\not U} \tag{6.15b}$$

Therefore,

$$F = F_U + F_{\not U} = \tfrac{1}{c^2\varepsilon}\mathcal{D}_U + \mu\mathcal{D}_{\not U} \tag{6.16}$$

Even more simply, we may write

$$\mu\mathcal{D}_U = n^2 F_U \tag{6.17a}$$

$$\mu \mathcal{D}_{\not{U}} = F_{\not{U}} \tag{6.17b}$$

where $n = c\sqrt{\mu\varepsilon}$ is the *proper index of refraction* of the material — that is, the index of refraction as measured in its rest frame. Observe that if $\varepsilon = \varepsilon_0$ and $\mu = \mu_0$, then $n = 1$ and $F = \mu_0 \mathcal{D}$.

6.1.4 Macroscopic Maxwell's Equation

As derived above, the complete set of Maxwell's equations in the macroscopic view of a dielectric material are given by

$$\Box \cdot \mathcal{D} = J \tag{6.18a}$$

$$\Box \wedge F = 0 \tag{6.18b}$$

We remind the reader once again that J, in this context, refers to the free four-current only; the inhomogeneous equation, originally equating the spacetime divergence of the Faraday field with the total charge and current, has been replaced by (6.18a) associating the displacement field with free charge and current. The homogeneous equation (6.18b), having no reference to sources of either kind (bound or free), is unchanged.

Analyses will be greatly simplified if we can unify these two equations into one as we did for the Faraday field in a vacuum. To that end, let us define a modified differential operator that scales the projection onto U (the four-velocity of the material in the observational reference frame) by the material's proper index of refraction,

$$\blacksquare = n\Box_U + \Box_{\not{U}} \tag{6.19}$$

Whereas the empty square represents the spacetime gradient in vacuum, I have used the filled square here to represent the spacetime gradient in a (homogeneous) region filled with material. In the rest frame of that material, this becomes

$$\blacksquare_{rest} = \sqrt{\mu\varepsilon}\gamma^0 \partial t + \gamma^1 \partial x + \gamma^2 \partial y + \gamma^3 \partial z \tag{6.20}$$

It is evident in this equation that the speed of light c has been replaced by $v_p = 1/\sqrt{\mu\varepsilon}$. We also define a macroscopic Faraday bivector field,

$$\mathcal{F} = \eta \left(n^{-1} \mathcal{D}_U + \mathcal{D}_{\not{U}} \right) = c \left(F_U + n^{-1} F_{\not{U}} \right) \tag{6.21}$$

where η is the proper wave impedance of the medium, or the wave impedance in its rest frame, given by

$$\eta = \sqrt{\frac{\mu}{\varepsilon}} \tag{6.22}$$

We use the calligraphic letter $\mathcal{F}$ to distinguish the effective macroscopic field from the vacuum Faraday field, F. The scaling factor has been selected so that the field quantities in the rest frame are those most commonly used in engineering,

$$\mathcal{F}_{rest} = \eta \left(n^{-1} c\mathbf{D} + i\mathbf{H}\right) = \mathbf{E} + i\eta\mathbf{H} \tag{6.23}$$

Equation (6.21) is easily inverted to find the microscopic (vacuum) Faraday field and displacement field,

$$\mathcal{D} = \eta^{-1} \left(n\mathcal{F}_U + \mathcal{F}_{\not U}\right) \tag{6.24a}$$

$$F = c^{-1} \left(\mathcal{F}_U + n\mathcal{F}_{\not U}\right) \tag{6.24b}$$

The spacetime divergence of the macroscopic Faraday field (in any reference frame) is given by

$$\blacksquare \cdot \mathcal{F} = \eta \left(n\square_U + \square_{\not U}\right) \cdot \left(n^{-1}\mathcal{D}_U + \mathcal{D}_{\not U}\right) \tag{6.25a}$$

$$= \eta \left(\square_U \cdot \mathcal{D}_U + n\square_U \cdot \mathcal{D}_{\not U} + n^{-1}\square_{\not U} \cdot \mathcal{D}_U + \square_{\not U} \cdot \mathcal{D}_{\not U}\right) \tag{6.25b}$$

$$= \eta \left(\square_U \cdot \mathcal{D}_U + n\square_U \cdot \mathcal{D}_{\not U} + n^{-1}\square_{\not U} \cdot \mathcal{D}_U + \square_{\not U} \cdot \left(\mathcal{D} - \mathcal{D}_U\right)\right) \tag{6.25c}$$

$$= \eta \left(\square_U \cdot \mathcal{D}_U + n\square_U \cdot \mathcal{D}_{\not U} + \left(n^{-1} - 1\right) \square_{\not U} \cdot \mathcal{D}_U + \square_{\not U} \cdot \mathcal{D}\right) \tag{6.25d}$$

Remember that the $\square$ operator behaves much like any other vector in dot and wedge products. This allows us to apply the projection identities in Section D.8.3 to some of the terms above. In particular, we may make the following substitutions,

$$\square_U \cdot \mathcal{D}_U = \square_U \cdot \mathcal{D} \tag{6.26a}$$

$$\square_U \cdot \mathcal{D}_{\not U} = 0 \tag{6.26b}$$

$$\square_{\not U} \cdot \mathcal{D}_U = \left(\square \cdot \mathcal{D}\right)_U \tag{6.26c}$$

Although strictly rigorous proofs require consideration of four-dimensional Minkowski spacetime, many of these identities can be visualized in the context of simple three-dimensional geometry, as shown in Figure 6.2. Applying these results to the field divergence formula, we have

$$\blacksquare \cdot \mathcal{F} = \eta \left(\square_U \cdot \mathcal{D} + \left(n^{-1} - 1\right) \left(\square \cdot \mathcal{D}\right)_U + \square_{\not U} \cdot \mathcal{D}\right) \tag{6.27a}$$

$$= \eta \left(\square \cdot \mathcal{D} + \left(n^{-1} - 1\right) \left(\square \cdot \mathcal{D}\right)_U\right) = \eta \left(J + \left(n^{-1} - 1\right) J_U\right) \tag{6.27b}$$

(a) (b)

Figure 6.2 Illustration of vector-projection product identities. (a) The tangential divergence of a bivector field projected onto the same vector as the derivative is equal to the tangential divergence of the field as a whole. (b) The tangential divergence of a bivector field rejected from the vector along which differentiation occurs is zero.

$$= \eta\left(n^{-1}J_U + J_{\not U}\right) = \mathcal{J} \tag{6.27c}$$

where $\mathcal{J}$ is to be recognized as the macroscopic source vector. A spacetime split in the rest frame provides some insight into this quantity,

$$\left(n^{-1}J_{c\gamma_0} + J_{\not{c\gamma_0}}\right)\gamma_0 = v_p\rho + \mathbf{J} \tag{6.28}$$

Note that this has the same form as the original four-current density except that the timelike component, the charge density, has an attached scaling factor of v_p instead of c.

Next, we consider the spacetime curl of the macroscopic Faraday field,

$$\blacksquare \wedge \mathcal{F} = v_p\left(n\square_U + \square_{\not U}\right) \wedge \left(nF_U + F_{\not U}\right) \tag{6.29a}$$

$$= v_p\left(n^2\square_U \wedge F_U + n\square_U \wedge F_{\not U} + n\square_{\not U} \wedge F_U + \square_{\not U} \wedge F_{\not U}\right) \tag{6.29b}$$

$$= v_p\left(n^2\square_U \wedge F_U + n\square_U \wedge F_{\not U} + n\square_{\not U} \wedge (F - F_{\not U}) + \square_{\not U} \wedge F_{\not U}\right) \tag{6.29c}$$

$$= v_p\left(n^2\square_U \wedge F_U + n\square_U \wedge F_{\not U} + n\square_{\not U} \wedge F + (1-n)\square_{\not U} \wedge F_{\not U}\right) \tag{6.29d}$$

Once again, we may consider the identities in Section D.8.3 to simplify this result. The following in particular are useful,

$$\square_U \wedge F_U = 0 \tag{6.30a}$$

$$\square_U \wedge F_{\not U} = \square_U \wedge F \tag{6.30b}$$

$$\Box_{\not{U}} \wedge F_{\not{U}} = (\Box \wedge F)_{\not{U}} \tag{6.30c}$$

Thus, we have

$$\blacksquare \wedge \mathcal{F} = v_p \left(n\Box_U \wedge F + n\Box_{\not{U}} \wedge F + (1-n)(\Box \wedge F)_{\not{U}} \right) \tag{6.31a}$$

$$= v_p \left(n\Box \wedge F + (1-n)(\Box \wedge F)_{\not{U}} \right) = 0 \tag{6.31b}$$

Finally, we put these two results together to form the single macroscopic Maxwell's equation,

$$\blacksquare\mathcal{F} = \mathcal{J} \tag{6.32}$$

For most cases that we will be concerned with, the free source vector, $\mathcal{J}$, will be zero.

6.1.5 Macroscopic Field Potential

It is also useful in some contexts to define a macroscopic field potential, $\mathcal{A}$, that has the same mathematical relationship to $\mathcal{F}$ as the microscopic potential A has to F,

$$\mathcal{F} = \blacksquare\mathcal{A} \tag{6.33}$$

under the material Lorenz gauge condition,

$$\blacksquare \cdot \mathcal{A} = 0 \tag{6.34}$$

Therefore,

$$\blacksquare\mathcal{F} = \blacksquare^2\mathcal{A} = \mathcal{J} \tag{6.35}$$

and $\mathcal{J}$ once again is the free source density.

6.1.6 Conducting Media

Recall that Ohm's law, or the proportional relationship between the electric field and current density in (1.40), is not a law, per se, rather it is a macroscopic, empirical observation explained by the relatively slow, field-induced drift of free charge carriers in the material added to the much more rapid and random thermal motions of those carriers which effectively dominate the mean time between collisions (see Section 1.4.2). Its generalization to a covariant form will thus likewise be an approximation, valid only within certain limits of velocity and field strength.

The random thermal motion of free, conduction electrons in metals at room temperature (and in their own rest frame) is around $10^5 - 10^6$ m/s, or about a thousandth the speed of light (decidedly nonrelativistic). The underlying mechanism of conductivity is thus bound to remain intact in most other reference frames, except that time dilation lengthens the mean time between collisions, giving an ambient field (in a moving reference frame) more opportunity to accelerate those charges. This translates to an increase in the apparent conductivity proportional to the Lorentz factor, $\cosh\zeta$.

Also, in the classical version of Ohm's law, the conductor is implicitly assumed to be at rest with respect to the observer, so that the random thermal motions of free charges in all directions average out to zero. Thus, any acceleration of charges due to an ambient magnetic field (which is dependent on the direction of motion of those charges) also averages to zero. If, instead, the conductor is allowed to be in motion, then the free charges contained within it will also have a net or average velocity in that direction, and some acceleration due to an ambient magnetic field may be observed,

$$\mathbf{J} = \sigma\left(\mathbf{E} + \mathbf{v}\times\mathbf{B}\right)\cosh\zeta \tag{6.36}$$

The expression in parentheses derives almost directly from (1.10), the Lorentz force, while the conductivity, σ, as before accounts for the dampening of the acceleration that this force would otherwise cause due to collisions with the lattice, modified by $\cosh\zeta$ due to dilation of the time interval between those collisions.

The covariant form of Ohm's law may therefore be inferred by analogy from the Lorentz force law previously derived in tensor form (4.102),

$$J^{\mu} = \sigma F^{\mu\nu}U_{\nu} \tag{6.37}$$

or, converting to a spacetime algebraic representation,

$$J = \sigma F\cdot U \tag{6.38}$$

Note that $J\cdot U = \sigma(F\cdot U)\cdot U = 0$; therefore, $J = J_{\not{U}}$.

To verify that we have written this correctly, we may perform a spacetime split on this equation,

$$J\gamma_0 = \sigma(F\cdot U)\gamma^0 = \tfrac{1}{2}\sigma(FU - UF)\gamma_0 \tag{6.39a}$$

$$\therefore c\rho + \mathbf{J} = \tfrac{1}{2}\sigma\left[\left(\tfrac{1}{c}\mathbf{E} + i\mathbf{B}\right)U - U\left(\tfrac{1}{c}\mathbf{E} + i\mathbf{B}\right)\right]\gamma_0 \tag{6.39b}$$

$$= \tfrac{1}{2}\sigma\left[\left(\tfrac{1}{c}\mathbf{E} + i\mathbf{B}\right)U\gamma_0 + U\gamma_0\left(\tfrac{1}{c}\mathbf{E} - i\mathbf{B}\right)\right] \tag{6.39c}$$

$$= \tfrac{1}{2}\sigma \left[\left(\tfrac{1}{c}\mathbf{E} + i\mathbf{B}\right)(c + \mathbf{v}) + (c + \mathbf{v})\left(\tfrac{1}{c}\mathbf{E} - i\mathbf{B}\right)\right] \cosh\zeta \tag{6.39d}$$

$$= \sigma \left[\tfrac{1}{2c}(\mathbf{vE} + \mathbf{Ev}) + \mathbf{E} - \tfrac{i}{2}(\mathbf{vB} - \mathbf{Bv})\right] \cosh\zeta \tag{6.39e}$$

$$= \sigma \left[\tfrac{1}{c}\mathbf{v} \cdot \mathbf{E} + \mathbf{E} + \mathbf{v} \times \mathbf{B})\right] \cosh\zeta \tag{6.39f}$$

This is a system of equations. We may isolate the scalar and vector parts,

$$c\rho = \tfrac{\sigma}{c}(\mathbf{v} \cdot \mathbf{E}) \cosh\zeta \tag{6.40a}$$

$$\mathbf{J} = \sigma(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cosh\zeta \tag{6.40b}$$

We see that (6.40b) does match (6.36). At rest ($\mathbf{v} = 0$ and $\cosh\zeta = 1$), this reduces to the classical form, $\mathbf{J} = \sigma\mathbf{E}$.

Equation (6.38) is clearly written for the fundamental, microscopic Faraday field. In principle, it applies to both free and bound sources, but then the conductivity of bound sources is zero by definition, at least at DC — minute AC displacement currents, rather than being associated with conductivity, are typically dealt with as a complex component of the dielectric constant. We may thus write the free source density in terms of the macroscopic Faraday field as follows,

$$\mathcal{J} = \eta\left(n^{-1}J_U + J_{\not{U}}\right) = \eta J_{\not{U}} = \eta\sigma F \cdot U = \tfrac{\eta\sigma}{c}\mathcal{F} \cdot U \tag{6.41}$$

A summary of the key results for fields in various media is given in Table 6.1.

6.2 WAVES IN MATTER

Failures to detect the motion of the luminiferous aether in the latter part of the nineteenth century led to various *aether-drag hypotheses* that proposed the aether is either fully or partially dragged along by matter in motion. Experiments designed to measure this dragging effect confirmed those hypotheses to some extent, but not precisely as their creators intended, and not without seeming to contradict one another [5].

Although belief in the aether is no longer necessary under special relativity, we do predict that matter, when present, retards the propagation of electromagnetic waves. Matter in motion, therefore, gives a boost to the retarded waves within it in the direction of motion, while further retarding those waves propagating against the motion of the medium.

Table 6.1
Equations for Fields in Media

Name	Separate Faraday and Displacement Fields (F,$\mathcal{D}$)	Combined Macroscopic Field ($\mathcal{F}$)
Constituent relationships	$F = \mu_0 \left(\mathcal{D} + \mathcal{M}\right)$ $\mathcal{D} = c^2 \varepsilon F_U + \frac{1}{\mu} F_{\not{U}}$	$\mathcal{F} = c\left(F_U + n^{-1} F_{\not{U}}\right)$ $= \eta\left(n^{-1}\mathcal{D}_U + \mathcal{D}_{\not{U}}\right)$
Source vector	$J = (c\rho + \mathbf{J})\gamma_0$	$\mathcal{J} = \eta\left(n^{-1} J_U + J_{\not{U}}\right)$
Differential operator	$\square = \gamma_0\left(\partial_0 + \boldsymbol{\nabla}\right)$	$\blacksquare = n\square_U + \square_{\not{U}}$
Maxwell's equation(s)	$\square \cdot \mathcal{D} = J$ $\square \wedge F = 0$	$\blacksquare \mathcal{F} = \mathcal{J}$
Ohm's law	$J = \sigma F \cdot U$	$\mathcal{J} = \frac{\eta\sigma}{c} \mathcal{F} \cdot U$
Field potentials (Lorenz gauge)	$F = \square A$	$\mathcal{F} = \blacksquare \mathcal{A}$
Lorenz gauge	$\square \cdot A = 0$	$\blacksquare \cdot \mathcal{A} = 0$
Potential wave equation	$\square^2 A = \mu_0\left(J_{free} + J_{bound}\right)$	$\blacksquare^2 \mathcal{A} = \mathcal{J}$

μ, ε, n, and η are the proper permeability, permittivity, index of refraction, and wave impedance.
J and $\mathcal{J}$ are the free (unbound) and macroscopic source densities, unless otherwise noted.

6.2.1 Plane Wave Solutions

Let us derive plane wave solutions in media as we did previously for plane waves in free space in Section 5.2.5. We will assume the medium has no free sources, so the macroscopic Maxwell's equation is

$$\blacksquare \mathcal{F} = 0 \tag{6.42}$$

As this is almost identical to the source-free form of Maxwell's equation in vacuum, our derivation will parallel that of Section 5.2.5 very closely. We will once again use the canonical polar form for the (macroscopic) Faraday field, $\mathcal{F}$, with a simple bivector prefactor, $\mathbf{f}$, that is constant throughout spacetime. Let us also this time combine two counter-rotating circularly polarized waves at the outset so that we may derive linearly polarized solutions,

$$\mathcal{F} = \tfrac{1}{2}\mathbf{f}\left(e^{i\varphi} + e^{-i\varphi}\right) = \mathbf{f}\cos\varphi \tag{6.43}$$

Therefore,

$$\blacksquare \mathcal{F} = -(\blacksquare\varphi)\mathbf{f}\sin\varphi = 0 \tag{6.44}$$

For the same reasons as before, we may conclude that both $\mathbf{f}$ and $\blacksquare\varphi$ are necessarily constant, share a common factor, and null in magnitude, but nonzero in value in order for us to arrive at nontrivial solutions. However, there is a difference this time in the relative scaling of temporal and spatial components within the differential operator, $\blacksquare$. This will lead to waves having velocity other than c. Thus, we may write

$$\mathbf{f} = sL \tag{6.45a}$$

$$\blacksquare\varphi = L \tag{6.45b}$$

$$\varphi = K \cdot x \tag{6.45c}$$

The original spacetime gradient, $\square$, ensured in our previous derivation that $L = K$ when the wave existed in a vacuum, but this time the two constants differ,

$$\blacksquare\varphi = \left(n\square_U + \square_{\not{U}}\right)(K \cdot x) = nK_U + K_{\not{U}} = L \tag{6.46}$$

Recall that the constant L is null. This equation therefore constrains K in a particular reference frame, with dependency on the motion of the material, U, and the direction of propagation of the wave with respect to that motion. To gain insight into that constraint, let us work a couple of examples.

6.2.2 Plane Waves in the Rest Frame

The simplest case corresponds to the rest frame of the material, wherein $U = c\gamma_0$. For simplicity, let us assume propagation in the γ_3 direction, with linear polarization in the γ_1 direction, as shown in Figure 6.3,

$$K = \tfrac{\omega}{c}\gamma_0 + k\gamma_3 \tag{6.47a}$$

$$s = \tfrac{1}{k}F_0\gamma_1 \tag{6.47b}$$

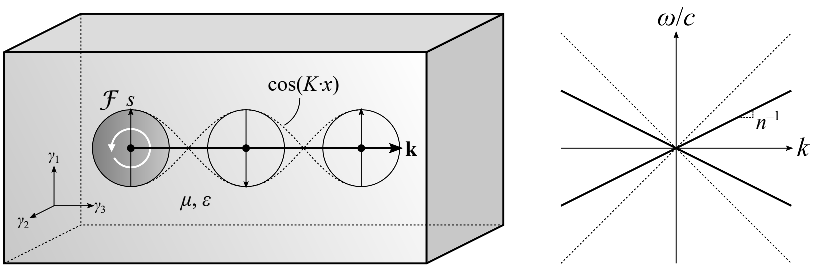
Therefore,

$$L = nK_U + K_{\not{U}} = nK_{c\gamma_0} + K_{\cancel{c\gamma_0}} = n\tfrac{\omega}{c}\gamma_0 + k\gamma_3 \tag{6.48a}$$

$$L^2 = \tfrac{n^2\omega^2}{c^2} - k^2 = 0 \tag{6.48b}$$

$$\therefore n\omega = kc \tag{6.48c}$$

$$\mathbf{f} = sL = \tfrac{1}{k}F_0\gamma_1\left(n\tfrac{\omega}{c}\gamma_0 + k\gamma_3\right) = F_0\left(\sigma_1 + i\sigma_2\right) \tag{6.48d}$$



(a) (b)

Figure 6.3 A linearly polarized plane wave in the rest frame of a material. (a) Illustration of the macroscopic Faraday field. (b) Dispersion diagram. The thick black lines indicate the range of possible wavevectors, and the slope of those lines indicate the propagation velocity.

where n, the proper index of refraction of the material, must be greater than or equal to 1. In all directions, then, the normalization of K is given by

$$K^2 = \tfrac{\omega^2}{c^2} - k^2 = -\tfrac{\omega^2}{c^2}\left(n^2 - 1\right) \leq 0 \tag{6.49}$$

The wavevector K is therefore spacelike, corresponding to slow waves traveling at speed less than c. The phase velocity is given by

$$v_p = \frac{\omega}{k} = \frac{c}{n} \tag{6.50}$$

If we associate σ_1, σ_2, and σ_3 with $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$, respectively, then the complete macroscopic Faraday field for a forward-traveling wave is

$$\mathcal{F} = \mathbf{f}\cos\varphi = F_0(\mathbf{x} + i\mathbf{y})\cos(\omega t - kz) \tag{6.51}$$

6.2.3 Longitudinal Waves in Moving Media

The next case we will consider is that of a moving medium with plane waves traveling along the same axis, either forward or backward. We could derive a solution from the general equations given in Section 6.2.1, but a simpler approach is to apply a Lorentz boost to the rest-frame solutions found in Section 6.2.2. A

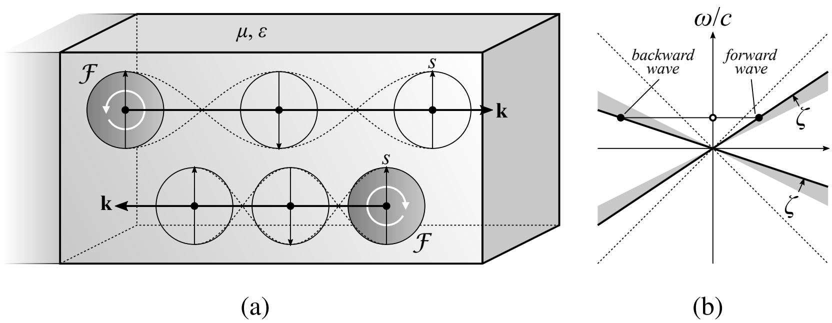


Figure 6.4 A linearly polarized plane wave in a longitudinally moving material. (a) Illustration of the macroscopic Faraday field. (b) Dispersion diagram showing nonreciprocity in the available wavevectors. The forward and backward-traveling waves shown have the same frequency, indicated by the intercepts to the horizontal line on the dispersion diagram.

basic consequence of special relativity is that the boosted velocities of objects and subluminal waves (such as those in matter) do not add linearly, but asymptotically approach c in the most rapidly countermoving inertial reference frames. Light waves in free space, already traveling at speed c, are, in effect, infinitely boosted, so that no further change in speed, either faster or slower, can be observed in any reference frame. Light waves in matter have been retarded by their interaction with the material constituents, so that boosts with respect to the rest frame of that medium do still result in some change in speed, faster if the wave is directed against the motion of the reference frame but with the apparent motion of the medium, and slower if the wave is directed with the motion of the reference frame and against the apparent motion of the medium.

Our goal is to derive wave solutions in a medium moving longitudinally, as shown in Figure 6.4(a). For simplicity, the positive reference direction for movement of the medium as seen by the observer is to be the same as for the forward traveling wave, in the $+z$ direction. Thus, we start with the solutions found in the rest frame and apply a reverse boost in the $-z$ direction. The appropriate rotor is

$$R = e^{-\zeta\gamma_{03}/2} \tag{6.52}$$

The wavevector then transforms as follows,

$$K' = RK\widetilde{R} = e^{-\zeta\gamma_{03}/2}\left(\tfrac{\omega}{c}\left(\gamma_0 + n\gamma_3\right)\right)e^{\zeta\gamma_{03}/2} \tag{6.53a}$$

$$= \tfrac{\omega}{c}\left[\left(\cosh\zeta + n\sinh\zeta\right)\gamma_0 + \left(n\cosh\zeta + \sinh\zeta\right)\gamma_3\right] = \tfrac{\omega'}{c}\gamma_0 + k'\gamma_3 \tag{6.53b}$$

$$\therefore v_p' = \frac{\omega'}{k'} = \frac{\cosh\zeta + n\sinh\zeta}{n\cosh\zeta + \sinh\zeta}c = \frac{1 + n\tanh\zeta}{n + \tanh\zeta}c \tag{6.53c}$$

If one substitutes $n = c/v_p$ above, we recover (3.4), the original velocity addition formula. Still, it is often a good idea to verify limiting cases when a formula such as this is derived. For very slow-moving media ($\tanh\zeta \approx 0$), we have $v_p' \approx c/n$, which confirms the definition of the index of refraction in the rest frame. For fast-moving media, approaching light speed ($\tanh\zeta \to 1$), we have $v_p' \to c$, which is consistent with the otherwise slow wave being dragged forward by the medium at nearly the speed of light. Finally, when there is no medium ($n = 1$), we have $v_p' = c$, as expected for an electromagnetic wave in free space.

A modified index of refraction can be defined,

$$n' = \frac{c}{v_p'} = \frac{n + \tanh\zeta}{1 + n\tanh\zeta} \tag{6.54}$$

What is most significant is that the index of refraction in this reference frame is asymmetric with respect to the direction of motion (the reverse direction would correspond to negative rapidity above). Motion of the medium has rendered it electrically *nonreciprocal*. This is indicated by the asymmetry of the wavevectors on the dispersion diagram of Figure 6.4(b). For a given frequency, the wavelength is lengthened for waves traveling with the medium, and shortened for waves traveling against it. Note that if the medium is moving fast enough, even backward-traveling waves would be dragged forward. In that extreme case, no backward-traveling waves would be permitted, only two different kinds of forward-traveling waves having different velocities.

The field itself can be transformed using (5.113), noting that both field components are perpendicular to the direction of the boost in this case,

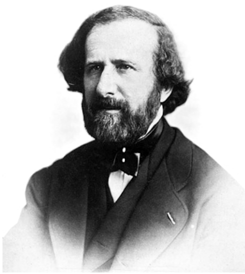
$$\mathbf{f}' = (\cosh\zeta - \gamma_{03}\sinh\zeta)\,\mathbf{f} = F_0\,(\cosh\zeta - \gamma_{03}\sinh\zeta)\,(\sigma_1 + i\sigma_2) \tag{6.55a}$$

$$= F_0(\cosh\zeta + \sinh\zeta)\,(\sigma_1 + i\sigma_2) \tag{6.55b}$$

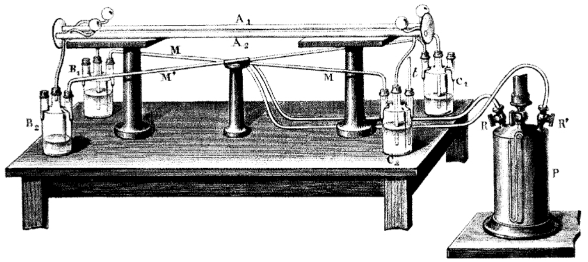
This represents an enhancement in the field amplitude when $\zeta > 0$ (the medium is dragging the wave forward), and an attenuation of the field amplitude otherwise.

6.2.4 Fizeau Water Experiment

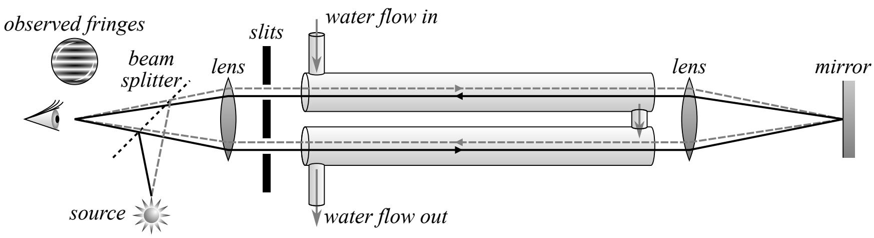
The results of the analysis in Section 6.2.3 are supported by a famous experiment conducted by Hippolyte Fizeau, Figure 6.5(a) (the same Fizeau whose direct mea-



(a)



(b)



(c)

Figure 6.5 (a) Hippolyte Fizeau. (b) Original sketch of the apparatus used in Fizeau's water experiment. (c) Experimental diagram. The light following the dashed gray path moves with the water, while the light following the solid black path moves against it.

surement of the speed of light in air was cited by Maxwell in his claim that electromagnetic waves and light were one and the same). Fizeau was an avid experimentalist who tested many properties of light, but the result of this experiment was of such profound importance to the development of special relativity that it is commonly known as the Fizeau experiment [6].

In Fizeau's time, it was thought that light was supported by and propagated through a luminiferous medium, the aether, and that the aether (rather than the light itself) could be dragged along by its contact with moving matter, such as water. Other experiments [7] had suggested that the aether was stationary with respect to the Earth, even as the Earth moved around the Sun, supporting the hypothesis that it was fully dragged along by the Earth's matter. Fizeau sought to confirm this by creating a situation in which the speed of light was measurably altered by its interaction with a moving medium.

Fizeau's apparatus, sketched in Figure 6.5(b) and drawn diagrammatically in Figure 6.5(c), directs the light from an external source through two slits and then through a long tube of flowing water, such that the light following one path is always moving against the flow of water, while the light following the other path is always moving with it. The two beams of light then recombine at the point of observation, creating a striped interference pattern known as fringes. Any mismatch in the delay between the two beams of light following different paths manifests by shifting the interference pattern, or fringes, away from the center.

If the aether was completely dragged along by the water, as prior experiments had suggested, then the total velocity of light in each path should simply be the velocity of light through the water, v_l, plus the velocity of the water, v_w. Thus, the total velocity in each of the two arms of the apparatus should be

$$v^+ = v_l + v_w \tag{6.56a}$$

$$v^- = v_l - v_w \tag{6.56b}$$

The displacement of the fringes should be proportional to the difference in these two velocities, or just $v^+ - v^- = 2v_w$. Instead, Fizeau measured a much smaller effect,

$$v^+ - v^- = 2v_w\left(1 - \frac{1}{n^2}\right) \tag{6.57}$$

where n as before is the index of refraction of the material, given by $n = c\sqrt{\varepsilon\mu}$. Since n is in general a function of frequency, this would imply that the aether could somehow simultaneously move at many different velocities (corresponding to the different wavelengths of light) at the same time.

We can now make sense of Fizeau's result in the context of special relativity. The velocity of light through the water in the water's rest frame is simply

$$v_l = \frac{c}{n} \tag{6.58}$$

just as Fizeau would expect. However, the total velocity in the laboratory frame is not given by the simple summations in (6.56), but rather by a Lorentz boost at velocity v_w. We may apply the velocity addition formula (3.4),

$$v^\pm = \frac{v_l \pm v_w}{1 \pm \frac{v_l v_w}{c^2}} = \frac{c \pm n v_w}{n \pm \frac{v_w}{c}} \tag{6.59a}$$

$$\therefore v^+ - v^- = \frac{c + n v_w}{n + \frac{v_w}{c}} - \frac{c - n v_w}{n - \frac{v_w}{c}} = 2v_w \frac{n^2 - 1}{n^2 - \left(\frac{v_w}{c}\right)^2} \tag{6.59b}$$

$$\approx 2v_w \left(1 - \frac{1}{n^2}\right) \tag{6.59c}$$

where the approximation holds when $v_w \ll c$. Fizeau's result is therefore merely a low-speed approximation of the relativistic velocity addition formula.

The index of refraction for visible light in water is about $n \approx 1.333$, so the deviation in differential velocity between the two paths in Fizeau's experiment from that predicted by nonrelativistic velocity addition is

$$1 - \frac{1}{n^2} = 1 - \frac{1}{1.333^2} \approx 0.44 \tag{6.60}$$

Take a moment to ponder the importance of this. The conventional, lazy argument against relativity having any practical value is that it is only significant at tremendous velocities and/or energy levels that are well beyond human capability to produce. Yet here is a very tangible result, the physical translation of wave interference patterns, brought about by a relativistic effect that alters the outcome by more than a factor of 2, in a compact apparatus comprising little more than a light source, some optics, and a weak pump. The tremendous velocity at play is merely that of the light itself through the water, at $0.75c$ ($= c/1.333$), well into the relativistic regime and ready to be modulated by the comparatively sluggish motion of the water.

As anyone schooled in microwave engineering knows, the exploitation of wave propagation and interference patterns for constructive purposes is already a common practice. All that is needed is a little creativity to turn this counterintuitive result to our advantage.

6.2.5 Off-Axis Waves in Moving Media

Let us now consider waves traveling in a direction that is not aligned with the motion of the medium. It will be easiest to start with the rest-frame solutions from Section 6.2.2 and apply a Lorentz boost. Note that an orthogonally directed boost does not result in a wave traveling orthogonal to the medium's motion; instead, an orthogonal component is added to the wave's forward velocity resulting in an angled path (think of a swimmer crossing a river being dragged downstream, thus following a diagonal path from the viewpoint of an observer on shore). Lorentz contraction and time dilation will further aberrate the path of the wave.

The most direct way to calculate these effects is to apply a Lorentz boost at an arbitrary angle, θ, with respect to the propagation of the wave in the rest frame, as shown in Figure 6.6(a). We start once again with a γ_3-directed wavevector in the

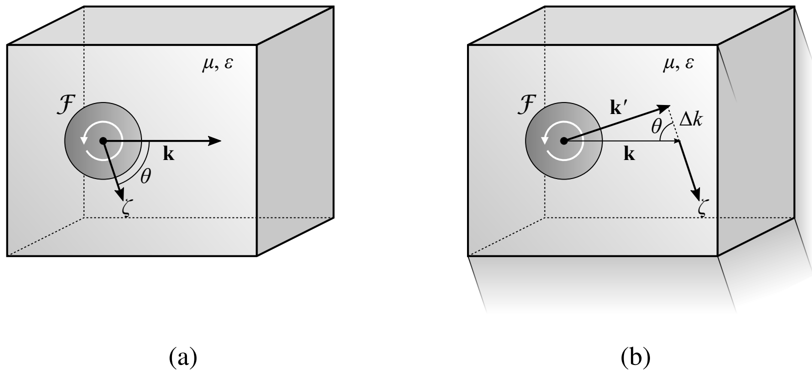


Figure 6.6 (a) A wave propagates in the rest frame of a material at an angle θ relative to the direction of a Lorentz boost ζ. (b) The wavenumber $\mathbf{k}$ is displaced by an amount Δk in a direction opposite to the boost.

rest frame,

$$K = \tfrac{\omega}{c}\gamma_0 + k\gamma_3 = \tfrac{\omega}{c}\left(\gamma_0 + n\gamma_3\right) \tag{6.61}$$

Next, we construct a rotor corresponding to the desired Lorentz boost at an angle θ from the γ_3 axis,

$$R = e^{\zeta G(\theta)/2} = \cosh\tfrac{\zeta}{2} + G(\theta)\sinh\tfrac{\zeta}{2} \tag{6.62}$$

where the generator of the rotation is

$$G(\theta) = \gamma_0\left(\gamma_3\cos\theta + \gamma_1\sin\theta\right) = \gamma_{03}\cos\theta + \gamma_{01}\sin\theta \tag{6.63}$$

Note that we have arbitrarily chosen a boost direction in the $\gamma_3 \wedge \gamma_1$ plane rather than the $\gamma_3 \wedge \gamma_2$ plane, or some other plane in between (it would have an impact on the transformed Faraday field, in consideration of how the wave in the rest frame is polarized, but for the wavevector alone it does not matter). The transformed wavevector is

$$K' = RK\widetilde{R} = e^{\zeta G(\theta)/2} K e^{-\zeta G(\theta)/2} \tag{6.64}$$

The original wavevector K has a γ_0 term and a γ_3 term. The γ_0 term anticommutes with the generator,

$$\gamma_0 e^{-\zeta G(\theta)/2} = e^{+\zeta G(\theta)/2}\gamma_0 \tag{6.65}$$

while the γ_3 term anticommutes only with the first part of the generator, $\gamma_{03}\cos\theta$, but commutes with the latter part, $\gamma_{01}\sin\theta$; therefore,

$$\gamma_3 e^{-\zeta G(\theta)/2} = \gamma_3 \left(\cosh\tfrac{\zeta}{2} - (\gamma_{03}\cos\theta + \gamma_{01}\sin\theta)\sinh\tfrac{\zeta}{2}\right) \tag{6.66a}$$

$$= \left(\cosh\tfrac{\zeta}{2} + (\gamma_{03}\cos\theta - \gamma_{01}\sin\theta)\sinh\tfrac{\zeta}{2}\right)\gamma_3 = e^{+\zeta G(-\theta)/2}\gamma_3 \tag{6.66b}$$

Putting these results together, we have

$$K' = \tfrac{\omega}{c} e^{\zeta G(\theta)/2}\left(\gamma_0 + n\gamma_3\right) e^{-\zeta G(\theta)/2} \tag{6.67a}$$

$$= \tfrac{\omega}{c}\left(e^{\zeta G(\theta)}\gamma_0 + n e^{\zeta G(\theta)/2} e^{\zeta G(-\theta)/2}\gamma_3\right) \tag{6.67b}$$

$G(\theta)$ and $G(-\theta)$ do not generally commute, so we cannot combine the exponents in the second term above using the product-of-powers rule (see Section 5.3.4). Instead, we must expand each rotor independently,

$$e^{\zeta G(\theta)/2} e^{\zeta G(-\theta)/2}$$

$$= \left(\cosh\tfrac{\zeta}{2} + G(\theta)\sinh\tfrac{\zeta}{2}\right)\left(\cosh\tfrac{\zeta}{2} + G(-\theta)\sinh\tfrac{\zeta}{2}\right) \tag{6.68a}$$

$$= \cosh^2\tfrac{\zeta}{2} + (G(\theta) + G(-\theta))\cosh\tfrac{\zeta}{2}\sinh\tfrac{\zeta}{2} + G(\theta)G(-\theta)\sinh^2\tfrac{\zeta}{2} \tag{6.68b}$$

$$= \cosh^2\tfrac{\zeta}{2} + 2\gamma_{03}\cos\theta\cosh\tfrac{\zeta}{2}\sinh\tfrac{\zeta}{2} + [\cos(2\theta) + \gamma_{31}\sin(2\theta)]\sinh^2\tfrac{\zeta}{2} \tag{6.68c}$$

$$= \tfrac{1}{2}\cosh\zeta + \tfrac{1}{2} + \gamma_{03}\cos\theta\sinh\zeta + \tfrac{1}{2}[\cos(2\theta) + \gamma_{31}\sin(2\theta)](\cosh\zeta - 1) \tag{6.68d}$$

Therefore,

$$K' = \tfrac{\omega}{c}\left(e^{\zeta G(\theta)}\gamma_0 + n e^{\zeta G(\theta)/2} e^{\zeta G(-\theta)/2}\gamma_3\right) \tag{6.69a}$$

$$\begin{aligned} &= \tfrac{\omega}{c}[\gamma_0\cosh\zeta - (\gamma_3\cos\theta + \gamma_1\sin\theta)\sinh\zeta] \\ &\quad + \tfrac{\omega n}{2c}[\gamma_3(\cosh\zeta + 1) - 2\gamma_0\cos\theta\sinh\zeta] \\ &\quad + \tfrac{\omega n}{2c}[\gamma_3\cos(2\theta) + \gamma_1\sin(2\theta)](\cosh\zeta - 1) \end{aligned} \tag{6.69b}$$

$$\begin{aligned} &= \tfrac{\omega}{c}\gamma_0(\cosh\zeta + n\cos\theta\sinh\zeta) \\ &\quad + \tfrac{\omega}{c}\gamma_1\left(\tfrac{n}{2}\sin(2\theta)(\cosh\zeta - 1) - \sin\theta\sinh\zeta\right) \\ &\quad + \tfrac{\omega}{c}\gamma_3\left(\tfrac{n}{2}(\cosh\zeta + 1) + \tfrac{n}{2}\cos(2\theta)(\cosh\zeta - 1) - \cos\theta\sinh\zeta\right) \end{aligned} \tag{6.69c}$$

$$= \tfrac{\omega}{c}\gamma_0\left(\cosh\zeta - n\cos\theta\sinh\zeta\right) + k\gamma_3$$
$$+ \tfrac{\omega}{c}\left(n\cos\theta(\cosh\zeta - 1) - \sinh\zeta\right)\left(\gamma_3\cos\theta + \gamma_1\sin\theta\right) \quad (6.69d)$$

To summarize, the boost transformation has yielded a new wavevector,

$$K' = \left(\tfrac{\omega'}{c} + \mathbf{k}'\right)\gamma_0 \quad (6.70)$$

having a modified frequency and wavenumber given by

$$\omega' = \omega(\cosh\zeta - n\cos\theta\sinh\zeta) \quad (6.71a)$$

$$\mathbf{k}' = \mathbf{k} - \Delta k\left(\sigma_3\cos\theta + \sigma_1\sin\theta\right) \quad (6.71b)$$

where

$$\Delta k = \tfrac{\omega}{c}\sinh\zeta - k\cos\theta(\cosh\zeta - 1) \quad (6.72)$$

Note that the displacement of the vector $\mathbf{k}$ is directly opposed to the boost direction, as shown in Figure 6.6(b). Furthermore, if we let $\theta = 0$ or π, then we have

$$\omega' = \omega(\cosh\zeta \pm n\sinh\zeta) \quad (6.73a)$$

$$k' = k \pm \Delta k = \tfrac{\omega}{c}\left(n\cosh\zeta \pm \sinh\zeta\right) \quad (6.73b)$$

in accord with the results of Section 6.2.3.

I have described the solutions in this section from the viewpoint of an initially forward traveling wave which is then redirected off-axis by a boost at an oblique angle. The resultant angle of propagation is not known a priori. The same results can be used to describe waves that do move in a prescribed direction, from the observer's standpoint, while the medium is in motion at an arbitrary angle.

First, we must identify in the solutions we have derived what is the relative angle between the motion of the medium and the consequent direction of propagation of the wave. From the observer's viewpoint, the motion of the medium after the boost is offset from the initial σ_3 axis by an angle $\pi - \theta$, where θ was the direction of the boost itself, as illustrated in Figure 6.7(a). The consequent direction of propagation of the wave is given by the spatial part of k', lying in the $\gamma_3 \wedge \gamma_1$ plane (the three-vector $\mathbf{k}'$ has σ_3 and σ_1 components). The angle of motion of the medium with respect to the wave is given by the difference between these two, as illustrated in Figure 6.7(b),

$$\phi = \pi - \theta - \tan^{-1}\left(\frac{\mathbf{k}'\cdot\sigma_1}{\mathbf{k}'\cdot\gamma_3}\right) = \pi - \theta - \tan^{-1}\left(\frac{-\Delta k\sin\theta}{k - \Delta k\cos\theta}\right) \quad (6.74a)$$

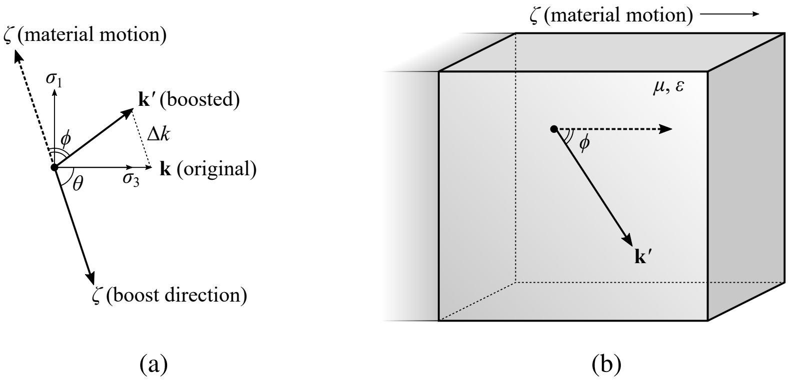


Figure 6.7 Determination of the propagation vectors in moving media. (a) Relationship between wavevectors in the rest frame and the observation frame. (b) Redrawn so that the reference direction for the motion of the material is to the right.

$$= \pi - \theta - \tan^{-1}\left(\frac{\sin\theta\sinh\zeta - n\sin\theta\cos\theta(\cosh\zeta - 1)}{\cos\theta\sinh\zeta - n\cos^2\theta(\cosh\zeta - 1) - n}\right) \tag{6.74b}$$

Next, we normalize out any transformation of frequency and wavenumber from these results by considering only their ratio, which is the velocity of propagation. The expressions are not simple, but easily tabulated, and can be used to create a three-dimensional contour, such as that drawn in Figure 6.8(a), representing the magnitude of velocity that a traveling wave can sustain through a moving medium in any direction. In this figure, the medium is moving to the right. Waves that are also propagating to the right are dragged forward, faster than they would be if the medium were at rest. Leftward traveling waves are dragged backwards, slowing them, and giving the velocity contour a flattened appearance on that side.

Since these results are independent of the polarization of the wave, each contour is a surface of revolution about the medium's axis of motion. It is therefore sufficient to draw only a cross-section, such as those in Figure 6.8(b). The innermost dark line corresponds to the medium at rest, and is a perfect circle since the material is isotropic, hence the wave velocity is the same in all directions. As the velocity of the medium increases, the left side of the circle flattens, while the right side expands. When the velocity of the medium becomes relativistic, even backward-traveling waves are dragged forward, which is why the contours fold back in on themselves around the origin. All contours must lie within the outermost dashed circle, corresponding to velocity c.

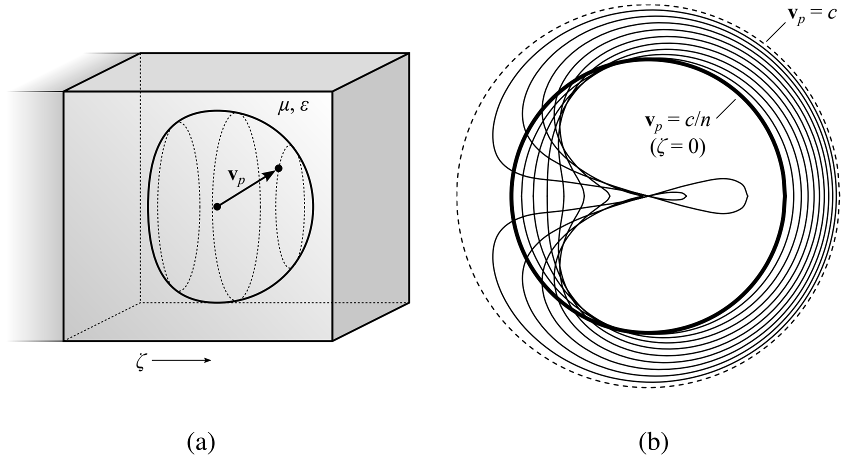


Figure 6.8 (a) Three-dimensional contour showing the velocity of propagation in any direction relative to the axis of motion of the medium. (b) Contour cross-sections for $n = 1.4$ and the rightward velocity of the medium varying from 0 to c in $0.1c$ steps. The thick, inner circle corresponds to the medium at rest, and the dashed outer circle corresponds to the vacuum speed of light, c.

6.2.6 Waves in Lossy Media

Thus far, we have only considered media that support propagating waves without losses. While this is a good approximation in most practical cases, no medium (except the vacuum of space, if that can be called a medium) is truly lossless. Losses come from many sources, but all lead to approximately the same behavior: an amplitude that decays exponentially with distance from the source. We shall specifically consider the case of a medium having finite conductivity, σ, as our representative model for this scenario.

We assume that our macroscopic Faraday field has the following form,

$$\mathcal{F} = \left(\mathbf{f}e^{i\varphi}\right) e^{\alpha} = \mathbf{f}e^{\alpha + i\varphi} \tag{6.75}$$

where $\mathbf{f}$ is constant throughout spacetime in accordance with the template for a plane wave, α is a scalar constant, and φ has the usual meaning as the invariant phase. Note that α and $i\varphi$ commute, so the product-of-powers rule applies.

We may write α and φ as the mapping of constant propagation vectors onto spacetime via the inner product,

$$\alpha = \mathcal{K} \cdot x \tag{6.76}$$

$$\varphi = K \cdot x \tag{6.77}$$

The spacetime gradient then yields the following auxiliary vectors,

$$\blacksquare\alpha = \mathcal{L} = n\mathcal{K}_U + \mathcal{K}_{\not{U}} \tag{6.78a}$$

$$\blacksquare\varphi = L = nK_U + K_{\not{U}} \tag{6.78b}$$

where U is the four-velocity of the medium.

Next, we substitute Ohm's law into Maxwell's equation in the presence of free sources,

$$\blacksquare\mathcal{F} = \mathcal{J} = \tfrac{\eta\sigma}{c}\mathcal{F} \cdot U \tag{6.79a}$$

$$(\blacksquare\alpha + \blacksquare\varphi i)\mathcal{F} = \tfrac{\eta\sigma}{c}\mathcal{F} \cdot U \tag{6.79b}$$

$$(\mathcal{L} + Li)\mathcal{F} + \tfrac{\eta\sigma}{c}U \cdot \mathcal{F} = 0 \tag{6.79c}$$

$$(\mathcal{L} + Li)\mathbf{f}e^{\alpha+i\varphi} + \tfrac{\eta\sigma}{c}U \cdot \left(\mathbf{f}e^{\alpha+i\varphi}\right) = 0 \tag{6.79d}$$

The scalar constant, e^{α}, is common to both terms and may be eliminated. We may also expand $e^{i\varphi}$ using Euler's formula. Since none of the other terms have any spacetime dependence, the equation can only be satisfied for all x if the terms in $\cos\varphi$ and $i\sin\varphi$ are satisfied independently,

$$(\mathcal{L} + Li)\mathbf{f}\cos\varphi + \tfrac{\eta\sigma}{c}(U \cdot \mathbf{f})\cos\varphi = 0 \tag{6.80a}$$

$$(\mathcal{L} + Li)\mathbf{f}i\sin\varphi + \tfrac{\eta\sigma}{c}(U \wedge \mathbf{f})i\sin\varphi = 0 \tag{6.80b}$$

After dropping the common trigonometric factors, we have

$$(\mathcal{L} + Li)\mathbf{f} + \tfrac{\eta\sigma}{c}U \cdot \mathbf{f} = 0 \tag{6.81a}$$

$$(\mathcal{L} + Li)\mathbf{f} + \tfrac{\eta\sigma}{c}U \wedge \mathbf{f} = 0 \tag{6.81b}$$

Now we add these two equations back together,

$$2(\mathcal{L} + Li)\mathbf{f} + \tfrac{\eta\sigma}{c}U\mathbf{f} = 0 \tag{6.82}$$

Recall from Section 6.2.1 that, in the lossless case, the phase gradient L was a null component of the bivector $\mathbf{f}$, such that

$$\mathbf{f} = sL \tag{6.83a}$$

$$\therefore L\mathbf{f} = 0 \tag{6.83b}$$

If we assume the conductivity is small, then these results should still hold, approximately. We are left with

$$\left(2\mathcal{L} + \tfrac{\eta\sigma}{c}U\right)\mathbf{f} = 0 \tag{6.84}$$

To satisfy this equation, we let the term in parentheses be colinear with the null factor of $\mathbf{f}$ (namely L),

$$2\mathcal{L} + \tfrac{\eta\sigma}{c}U = \lambda L \tag{6.85}$$

or, in the rest frame,

$$2\mathcal{L} + \eta\sigma\gamma_0 = \lambda L \tag{6.86}$$

Under a spacetime split, we expect L and $\mathcal{L}$ to have the following forms,

$$L\gamma_0 = \tfrac{n\omega}{c} + \mathbf{k} = k\left(1 + \hat{\mathbf{k}}\right) \tag{6.87a}$$

$$\mathcal{L}\gamma_0 = \alpha_c\hat{\mathbf{k}} \tag{6.87b}$$

where $\hat{\mathbf{k}}$ is the unit vector in the direction of propagation. The second equation shows that the direction of amplitude decay, $\mathcal{L}$, is the same as the direction of phase rotation, $\hat{\mathbf{k}}$, but with magnitude α_c, the attenuation constant for conductivity. Note that we do not expect $\mathcal{L}$ to have any temporal component as this would imply a wave source that grows or weakens over time, independent of space (although such a scenario is in no way forbidden by this derivation). Under the assumption of a constant amplitude source, then, we find

$$\eta\sigma = \lambda k \tag{6.88a}$$

$$\therefore \alpha_c = \tfrac{1}{2}\eta\sigma \tag{6.88b}$$

in agreement with the attenuation constant for good dielectrics derived via classical electromagnetics [3].

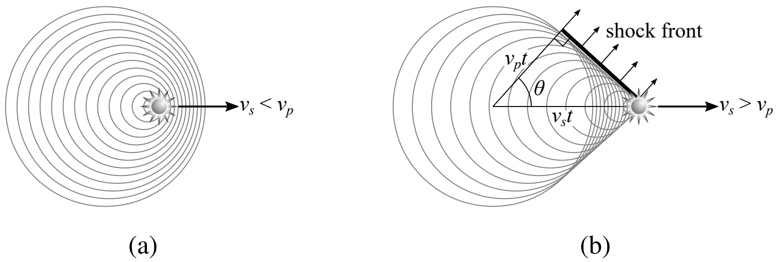


Figure 6.9 Čerenkov radiation. (a) A source moving slower than the phase velocity emits Doppler-shifted waves. (b) A source moving faster than the phase velocity emits a shock wave known as Čerenkov radiation.

6.2.7 Čerenkov Radiation

A final characteristic of Faraday waves propagating in bulk material that we will discuss is known as *Čerenkov radiation* [8, 9]. Although a detailed analysis is beyond the scope of this book, it deserves at least a brief mention, as it only occurs when charges move at relativistic speeds.

Nevertheless, the origin of Čerenkov radiation has nothing specifically to do with relativity. It is not so much a unique property of light as it is the electromagnetic manifestation of the same general behavior exhibited by all kind of waves when emitted by a source traveling more rapidly than the waves themselves. The most familiar nonelectromagnetic example is the sonic boom created by an aircraft traveling faster than the speed of sound. While special relativity precludes the possibility of any source traveling faster than the speed of light in vacuum (c), it is entirely possible for a source of electromagnetic energy, say a charged particle like an electron, to move faster through a material than the phase velocity of light in that material, $v_p = c/n$, if sufficiently energized (Čerenkov radiation is responsible for the characteristic blue glow surrounding underwater nuclear reactor cores).

The basic concept of Čerenkov radiation (and of sonic booms) is illustrated in Figure 6.9. In the first graphic, Figure 6.9(a), the source is moving with speed less than the phase velocity of the medium. While there is a Doppler effect causing the waves issued forward of the source to be perceived at a higher frequency, the wave train is still distributed from the viewpoint of a stationary observer. In Figure 6.9(b), because the source velocity exceeds the phase velocity, there are no waves ahead of the source, rather the source leaves trailing behind it a cone-shaped shock front where all of its past emitted wavefronts have piled up. This appears to a stationary observer as an intense, directional burst of energy when the shock wave passes.

The trailing angle of the cone-shaped shock wave is easily deduced from the proper index of refraction and the velocity of the source. Consider an interval of time, t. Over that interval, the source has moved by a distance

$$v_s t = \beta c t \tag{6.89}$$

where v_s is the velocity of the source and β is the associated Jackson number. The wavefront emitted at the beginning of the interval has travelled a shorter distance given by

$$v_p t = \frac{ct}{n} \tag{6.90}$$

The angle of the emission is then

$$\cos\theta = \frac{v_p t}{v_s t} = \frac{ct/n}{\beta c t} = \frac{1}{n\beta} \tag{6.91}$$

6.3 MATERIAL INTERFACES

It is common practice in engineering to manipulate the behavior of waves not only by selecting the bulk properties of the material through which the waves are propagating, but also by setting up the appropriate boundary conditions to guide their movements. It is therefore instructive to build upon what we know about material interactions to study how the field bivectors behave asymptotically at a material interface.

6.3.1 Geometry of Planar Surfaces

It is a virtual certainty that anyone reading this book will already know how the normal and tangential components of the electromagnetic field behave at the boundary between two media. The point is not to derive new results, rather it is to derive the familiar results in a novel way, and hopefully to gain new insights from the derivation. Prerequisite to attempting such a derivation in the present context of material interfaces is how one defines a (locally) planar boundary in four-dimensional spacetime in the first place.

In three dimensions, the orientation of a spatial plane is traditionally described by its surface normal, $\mathbf{n}$. We may suppose that this is the three-vector component of a spacetime split applied to the normal four-vector, N,

$$N\gamma_0 = N^0 + \mathbf{n} \tag{6.92}$$

where the temporal component, N^0, is zero in the rest frame of the surface. There is some validity to this approach, as we shall see, but spacetime algebra furnishes us with a richer toolset than that. We already know that blades of grade k are representative of k-dimensional geometries in spacetime. Might, then, the set of points residing on a two-dimensional planar surface be represented more directly as a grade-2 bivector than as the set of points orthogonal to its surface normal?

Not quite. Recall that the spatial coordinates themselves in four-dimensional spacetime are represented as bivector planes, as was illustrated in Figure 5.6. Spatial axis $\mathbf{x}$ is not merely associated with basis vector γ_1, rather it is the vector γ_1 for all times on γ_0, hence the bivector $\mathbf{x} = \gamma_1 \wedge \gamma_0 = \sigma_1$. To put it another way, a spacetime event identified by coordinates (x_0, x_1, x_2, x_3) occurs on the $\mathbf{x} = \sigma_1$ axis if and only if $x_2 = x_3 = 0$, regardless of the values of x_0 and x_1.

Similarly, the $\mathbf{xy}$ plane is not associated solely with the bivector γ_{12}, but with the bivector γ_{12} for all times on γ_0. The aforementioned spacetime event occurs on the $\mathbf{xy}$ plane if and only if $x_3 = 0$, regardless of the values of x_0, x_1, and x_2. We see, then, that a spatial planar surface is not described directly by a grade-2 bivector, but a grade-3 trivector or pseudovector, $P = \langle P \rangle_3$.

The normal four-vector that we alluded to earlier is in fact perpendicular to this pseudovector,

$$N \cdot P = 0 \tag{6.93}$$

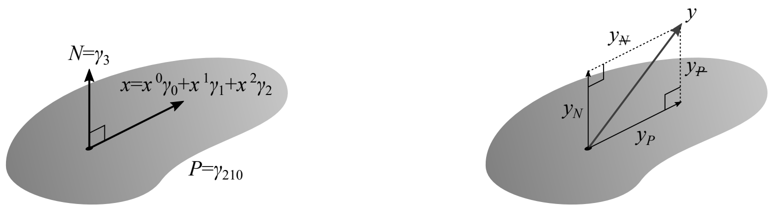
Moreover, P is the Hodge dual of N,

$$P = Ni^{-1} = iN \tag{6.94}$$

Take, for example, the $\mathbf{xy}$ plane, assumed to be at rest in a given reference frame. Points on the surface are contained for all times on γ_0 in the bivector γ_{21}, which is the pseudovector $P = \gamma_{210}$, as shown in Figure 6.10(a). The surface normal is then $N = Pi = \gamma_{210}i = \gamma_3$.

We typically enforce the condition that both the normal four-vector and surface pseudovector have unit magnitude,

$$|N|^2 = |P|^2 = 1 \tag{6.95}$$



(a) (b)

Figure 6.10 (a) A spatial planar surface defined by a normal four-vector (e.g., $N = \gamma_3$) or a pseudovector (e.g., $P = \gamma_{210}$). (b) A spacetime point y can be written equivalently as the sum of its projections onto P and N, or as the sum of its rejections from P and N, or as the sum of both its projection and rejection of either P or N. Points that lie within the plane have no projection onto N nor any rejection from P.

However, as the plane is spacelike, both elements have a negative signature in the metric we have chosen,

$$N^2 = P^2 = -1 \tag{6.96}$$

This ensures that $N^{-1} = -N$ and $P^{-1} = -P$. Points that lie on the plane are characterized by having no projection onto N, and no rejection from P,

$$x_N = 0 \tag{6.97}$$

$$x_{\not P} = 0 \tag{6.98}$$

Therefore,

$$x \cdot N = 0 \tag{6.99}$$

$$x \wedge P = 0 \tag{6.100}$$

Furthermore, the projection of x onto P or the rejection of x from N returns x itself, unchanged,

$$x_P = x \tag{6.101a}$$

$$x_{\not N} = x \tag{6.101b}$$

Note that any vector, y, can be expanded in terms of the blades P and N, as though written in a new vector basis,

$$y = y_P + y_{\not P} = y_P + y_N \tag{6.102}$$

This is illustrated in Figure 6.10(b).

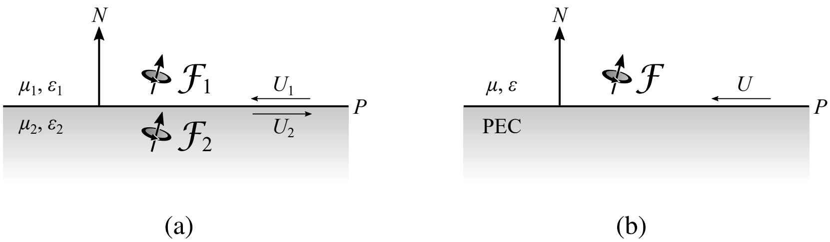


Figure 6.11 Interface between (a) two dielectric materials, and (b) a dielectric material and a perfect electric conductor.

6.3.2 Dielectric Boundaries

Let us now consider the boundary between two dielectric materials, one having proper constituent parameters μ_1 and ε_1 and the other having proper constituent parameters μ_2 and ε_2. Moreover, let the surface be defined by a spacelike unit normal vector N and pseudovector P, as shown in Figure 6.11(a).

Dielectric media are typically nonconducting, or very nearly so. We therefore assume that there is no free charge or current in either medium near this interface. However, the source-free macroscopic form of Maxwell's equation ($\blacksquare\mathcal{F} = 0$) is difficult to apply in this situation because the differential operator, $\blacksquare$, is defined in a way that depends on the material constituent parameters, and is better used in homogeneous dielectric regions (throughout which the constituent parameters are constant). In this case, the operator itself changes from one side of the boundary to the other. Instead, let us use the microscopic form of the equations and then recast those results in terms of the macroscopic Faraday field using (6.24) when needed.

The microscopic, inhomogeneous Maxwell's equation, or Gauss-Ampère law, in a source-free region reduces to

$$\square \cdot \mathcal{D} = (\square_P + \square_N) \cdot \mathcal{D} = 0 \tag{6.103}$$

where the spacetime gradient has been expanded via projection into components which are alternately tangential to the surface and normal to it.

Since the constituent parameters change discontinuously at the surface, the displacement field may likewise change discontinuously over an arbitrarily small distance in the N direction, while all other derivatives in both space and time are smooth. Thus, asymptotically, the derivative crossing the surface ($\square_N$) dominates while the derivative along the surface ($\square_P$) diminishes in significance. To put it another way, the derivative equation (6.103) can only be satisfied if the normal component of $\mathcal{D}$ is the same on both sides of the boundary

$$\mathcal{D}_1 \cdot N = \mathcal{D}_2 \cdot N \tag{6.104}$$

Otherwise, the term $\Box_N \cdot \mathcal{D}$ would diverge to infinity. This condition can equivalently be written

$$\mathcal{D}_1 \wedge P = \mathcal{D}_2 \wedge P \tag{6.105}$$

It is important to note that these boundary conditions apply in any reference frame, regardless of the apparent motion of either medium, U_1 or U_2, in that frame. The field quantities in a particular reference frame defined by the worldline γ_0, and where the boundary is static ($N^0 = 0$), may be determined using a spacetime split,

$$(\mathcal{D}_1 \cdot N)\,\gamma_0 = (\mathcal{D}_2 \cdot N)\,\gamma_0 \tag{6.106a}$$

$$((c\mathbf{D}_1 + i\mathbf{H}_1) \cdot N)\,\gamma_0 = ((c\mathbf{D}_2 + i\mathbf{H}_2) \cdot N)\,\gamma_0 \tag{6.106b}$$

$$c\mathbf{n} \cdot \mathbf{D}_1 + \mathbf{n} \times \mathbf{H}_1 = c\mathbf{n} \cdot \mathbf{D}_2 + \mathbf{n} \times \mathbf{H}_2 \tag{6.106c}$$

Therefore, after separating terms by grade,

$$\mathbf{n} \cdot \mathbf{D}_1 = \mathbf{n} \cdot \mathbf{D}_2 \tag{6.107a}$$

$$\mathbf{n} \times \mathbf{H}_1 = \mathbf{n} \times \mathbf{H}_2 \tag{6.107b}$$

Simply put, the normal component of $\mathbf{D}$ and the tangential components of $\mathbf{H}$ are continuous across a dielectric boundary. This result should be familiar to those versed in classical electromagnetic theory.

A similar analysis starting with the homogeneous Maxwell's equation ($\Box \wedge F = 0$) yields the following two equivalent forms,

$$F_1 \wedge N = F_2 \wedge N \tag{6.108a}$$

$$F_1 \cdot P = F_2 \cdot P \tag{6.108b}$$

or, in a given reference frame,

$$\mathbf{n} \times \mathbf{E}_1 = \mathbf{n} \times \mathbf{E}_2 \tag{6.109a}$$

$$\mathbf{n} \cdot \mathbf{B}_1 = \mathbf{n} \cdot \mathbf{B}_2 \tag{6.109b}$$

Thus, the tangential components of $\mathbf{E}$ and the normal component of $\mathbf{B}$ are also continuous across a dielectric boundary.

6.3.3 Conducting Boundaries

Of particular importance in electrical engineering is the behavior of fields at the boundary of a good conductor, as in Figure 6.11(b). In a *perfect electric conductor* (PEC), the conductivity is assumed to be infinite. According to Ohm's law, then, the Faraday fields inside such a conductor must be zero, lest the four-current also become infinite.

In this case since the only dielectric region is homogeneous, it is reasonable to start with the macroscopic field equation,

$$\blacksquare \mathcal{F} = \mathcal{J} \tag{6.110}$$

As we know, this is a mixed-grade equation that can be separated into two parts,

$$\blacksquare \cdot \mathcal{F} = \mathcal{J} \tag{6.111a}$$

$$\blacksquare \wedge \mathcal{F} = 0 \tag{6.111b}$$

Arguing in the same manner as before, we find that the asymptotic behavior of these equations near the boundary at which the fields have zero value inside the conductor yields the following results,

$$\mathcal{F} \cdot N = \mathcal{J}_s \tag{6.112a}$$

$$\mathcal{F} \wedge N = 0 \tag{6.112b}$$

where $\mathcal{J}_s$ represents the (free) source surface density (as opposed to the usual volume density). In expanded form for the rest frame of the dielectric,

$$(\mathcal{F} \cdot N)\gamma_0 = \eta \left(n^{-1} \left(J_s \right)_{c\gamma_0} + \left(J_s \right)_{\cancel{c\gamma_0}} \right) \gamma_0 \tag{6.113a}$$

$$\mathbf{n} \cdot \mathbf{E} + \eta \mathbf{n} \times \mathbf{H} = \eta \left(n^{-1} c\rho_s + \mathbf{J}_s \right) \tag{6.113b}$$

Therefore, separating by grade,

$$\mathbf{n} \cdot \mathbf{D} = \rho_s \tag{6.114a}$$

$$\mathbf{n} \times \mathbf{H} = \mathbf{J}_s \tag{6.114b}$$

and

$$(\mathcal{F} \wedge N)\gamma_0 i = -\eta \mathbf{n} \cdot \mathbf{H} + \mathbf{n} \times \mathbf{E} = 0 \tag{6.115}$$

so that

$$\mathbf{n} \cdot \mathbf{H} = \mathbf{n} \cdot \mathbf{B} = 0 \tag{6.116a}$$

$$\mathbf{n} \times \mathbf{E} = \mathbf{n} \times \mathbf{D} = 0 \tag{6.116b}$$

For convenience, let us restate (6.112) using the properties of projection and rejection,

$$\mathcal{F}N = \mathcal{J}_s \tag{6.117a}$$

$$\mathcal{F}_{\not{N}} = 0 \tag{6.117b}$$

The first of these can be used to determine the surface charge and current when a field solution is already known, but it is not a useful way to constrain a potential field solution for which the surface charge and current density are not known a priori. In that case, a differential form may help. Consider that because $\mathcal{F}_{\not{N}} = 0$, we have $\mathcal{F} = \mathcal{F}_N$ and we may write Maxwell's equation above the surface of the conductor (where there are no sources) as follows,

$$\blacksquare\mathcal{F} = \blacksquare\mathcal{F}_N = 0 \tag{6.118}$$

Taking the dot product of this equation with N and applying the identity (D.141), we find

$$N \cdot (\blacksquare\mathcal{F}_N) = (N \cdot \blacksquare)\mathcal{F}_N - \blacksquare\,(N \cdot \mathcal{F}_N) = 0 \tag{6.119}$$

Multiplying now by N^{-1} completes the projection of both $\blacksquare$ and $\mathcal{F}_N$, but since the latter is already projected onto N, this has no effect,

$$\blacksquare_N\mathcal{F}_N - \blacksquare\mathcal{F}_N = 0 \tag{6.120}$$

Finally, we recognize that the second term is already known to be zero; therefore,

$$\blacksquare_N\mathcal{F}_N = \blacksquare_N\mathcal{F} = 0 \tag{6.121}$$

Whereas (6.117a) constrains the value of the Faraday field at the surface of a conductor, (6.121) constrains its slope.

A summary of the boundary conditions for dielectric and conducting media is given in Table 6.2.

Table 6.2

Electromagnetic Boundary Conditions

Boundary Type	Classical Equations	Spacetime Calculus
Dielectric	$\mathbf{n} \times (\mathbf{H}_2 - \mathbf{H}_1) = 0$ $\mathbf{n} \cdot (\mathbf{D}_2 - \mathbf{D}_1) = 0$ $\mathbf{n} \times (\mathbf{E}_2 - \mathbf{E}_1) = 0$ $\mathbf{n} \cdot (\mathbf{B}_2 - \mathbf{B}_1) = 0$	$(\mathcal{D}_2 - \mathcal{D}_1) \cdot N = (\mathcal{D}_2 - \mathcal{D}_1) \wedge P = 0$ $(F_2 - F_1) \wedge N = (F_2 - F_1) \cdot P = 0$
PEC*	$\mathbf{n} \times \mathbf{H} = \mathbf{J}_s$ $\mathbf{n} \cdot \mathbf{D} = \rho_s$ $\mathbf{n} \times \mathbf{E} = 0$ $\mathbf{n} \cdot \mathbf{B} = 0$ $\frac{\partial \mathbf{H}}{\partial n} = 0$	$\mathcal{F} N = \mathcal{J}_s$ $\mathcal{F} \wedge N = 0$ $\blacksquare_N \mathcal{F} = 0$

*U is the four-velocity of the dielectric region in the observational reference frame.

6.4 WAVE REFLECTION AND REFRACTION

In addition to the asymptotic properties of the Faraday field at material interfaces, we are interested in how propagating waves interact with those boundaries. Classically, we envision an incident wave impacting on such a boundary, launching a reflected wave back into the source medium and a refracted or transmitted wave into the target medium, as shown in Figure 6.12.

6.4.1 Reflection from a Conducting Surface

If the target medium is a perfect conductor, then the field inside it must be zero, precluding the possibility of a refracted wave. We thus have only the incident and reflected waves. We may write the total macroscopic Faraday field above the surface of the source medium as the sum of these two plane-wave contributions,

$$\mathcal{F} = \mathbf{f}_i \cos(K_i \cdot x) + \mathbf{f}_r \cos(K_r \cdot x) \tag{6.122}$$

To this, we may apply the boundary condition from Table 6.2,

$$\mathcal{F} \wedge N = \mathbf{f}_i \wedge N \cos(K_i \cdot x) + \mathbf{f}_r \wedge N \cos(K_r \cdot x) = 0 \tag{6.123}$$

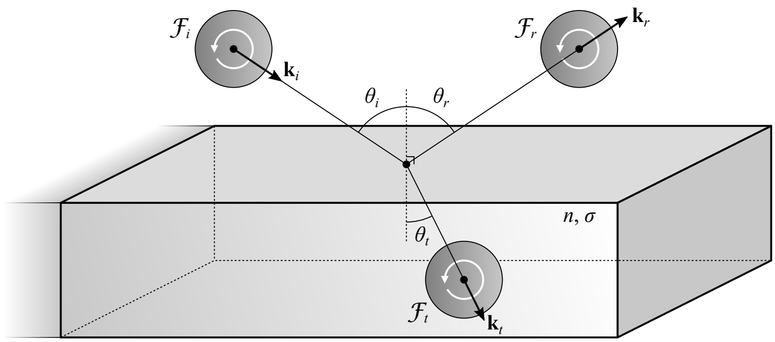


Figure 6.12 A propagating wave incident from one material or medium upon another. Plane-wave components in the source medium include incident and reflected waves, and those in the target medium comprise a refracted or transmitted wave. For simplicity, n is the relative (proper) index of refraction of the target medium.

Since both $\mathbf{f}_i$ and $\mathbf{f}_r$ are constant, this equation can only be satisfied for all points x that lie on the plane when

$$\mathbf{f}_i = -\mathbf{f}_r \tag{6.124a}$$

$$K_i \cdot x = K_r \cdot x \tag{6.124b}$$

Let us define $\Delta K = K_r - K_i$. To enforce that x lies in the plane of the interface in (6.124b), we may substitute from (6.101b),

$$\Delta K \cdot x = \Delta K \cdot x_{\not{N}} = 0 \tag{6.125a}$$

$$\Delta K \cdot \big(N^{-1}(N \wedge x)\big) = 0 \tag{6.125b}$$

$$(\Delta K \cdot N)(N \wedge x) - N\left(\Delta K \cdot (N \wedge x)\right) = 0 \tag{6.125c}$$

where we have used the fact that $N^{-1} = -N$. We now multiply both sides by N,

$$(\Delta K \cdot N)N \cdot (N \wedge x) + \Delta K \cdot (N \wedge x) = 0 \tag{6.126a}$$

$$((\Delta K \cdot N)N + \Delta K) \cdot (N \wedge x) = 0 \tag{6.126b}$$

Because this must be satisfied for all x, we may drop the second term $(N \wedge x)$,

$$\therefore \Delta K = -(\Delta K \cdot N)N = (\Delta K \cdot N)N^{-1} = (\Delta K)_N \tag{6.127}$$

showing that ΔK is equal to its own projection onto N; more succinctly, $K_r = K_i + \alpha N$. Since the incident and reflected waves lie in the same medium, K_i and K_r must have the same magnitude, and the angle of incidence is equal to the angle of reflection ($\theta_i = \theta_r$ in Figure 6.12). This is the well-known law of specular reflection. Although those angles vary from one reference frame to the next, their equality, the law of specular reflection, is frame-independent.

6.4.2 Refraction into a Secondary Dielectric

If the target medium is not a conductor, but another dielectric, then a refracted wave transmitted into the target medium must also be considered. Here we must recognize the difference between a *static* or *comoving interface*, wherein a reference frame exists for which both materials are at rest, and a *dynamic interface*, wherein one material is sliding across the other, as in a fluid against the surface of a container. In the former case, we may use the proper indices of refraction of both materials in the common rest frame; but in the latter, only one material can be at rest in any given reference frame. The effective index of refraction of the material in motion is direction-dependent, in accordance with the results of Section 6.2.5.

This time we shall use the microscopic field equations. The fields within the source medium are given by

$$F_1 = \mathbf{f}_i \cos(K_i \cdot x) + \mathbf{f}_r \cos(K_r \cdot x) \tag{6.128a}$$

$$\mathcal{D}_1 = \mathbf{d}_i \cos(K_i \cdot x) + \mathbf{d}_r \cos(K_r \cdot x) \tag{6.128b}$$

and those within the target medium are

$$F_2 = \mathbf{f}_t \cos(K_t \cdot x) \tag{6.129a}$$

$$\mathcal{D}_2 = \mathbf{d}_t \cos(K_t \cdot x) \tag{6.129b}$$

Applying now the boundary conditions for a dielectric interface,

$$\mathbf{f}_i \wedge N \cos(K_i \cdot x) + \mathbf{f}_r \wedge N \cos(K_r \cdot x) = \mathbf{f}_t \wedge N \cos(K_t \cdot x) \tag{6.130a}$$

$$\mathbf{d}_i \cdot N \cos(K_i \cdot x) + \mathbf{d}_r \cdot N \cos(K_r \cdot x) = \mathbf{d}_t \cdot N \cos(K_t \cdot x) \tag{6.130b}$$

As before, we require that all contributing phase terms are equal in the plane of the interface,

$$K_i \cdot x = K_r \cdot x = K_t \cdot x \tag{6.131}$$

We still have K_i and K_r in the same material, equal in magnitude, and thus making equivalent angles (specular reflection) with respect to the surface. For the transmitted wave, the same derivation as before leads us to the condition,

$$K_t = K_i + \alpha N \tag{6.132}$$

A spacetime split can be used to parse out the frequency and wavenumber into separate equations,

$$\frac{\omega_t}{c} = \frac{\omega_i}{c} \tag{6.133a}$$

$$\mathbf{k}_t = \mathbf{k}_i + \alpha \mathbf{n} \tag{6.133b}$$

where we have recalled that $N^0 = 0$ in the rest frame of either material (it does not matter which, for any residual motion of one material with respect to the other, when the first is at rest, must be perpendicular to the interface). The first, scalar equation gives us the fairly obvious result that the frequency of both waves is the same. Taking the cross product of the second, three-dimensional vector equation with the surface normal, $\mathbf{n}$, we have

$$\mathbf{n} \times \mathbf{k}_t = \mathbf{n} \times \mathbf{k}_i + \alpha \mathbf{n} \times \mathbf{n} = \mathbf{n} \times \mathbf{k}_i \tag{6.134a}$$

$$k_t \sin\theta_t = k_i \sin\theta_i \tag{6.134b}$$

Graphically, this means that the incident and transmitted wavevectors have the same extent lateral to the surface, as illustrated in Figure 6.13.

If the materials at the interface are comoving, then we can assume the common rest frame and apply the proper indices of refraction to k_t and k_i,

$$\therefore \frac{\sin\theta_i}{\sin\theta_t} = \frac{k_t}{k_i} = \frac{n_2}{n_1} = n \tag{6.135}$$

where n is the relative (proper) index of refraction between the target material and the source material. This we may recognize as *Snell's law of refraction*.

More generally, the target material is in motion relative to the first, and the effective index of refraction in the observational reference frame depends on the rapidity of movement as well as propagation angle. The refractive angle can be found graphically using the velocity contours derived earlier in Figure 6.8. The radius of the contour for index of refraction is the inverse of that for velocity relative to c. Thus, the index of refraction is larger in directions opposite to the motion of the material, as shown in the lower hemisphere of the diagram in Figure 6.13.

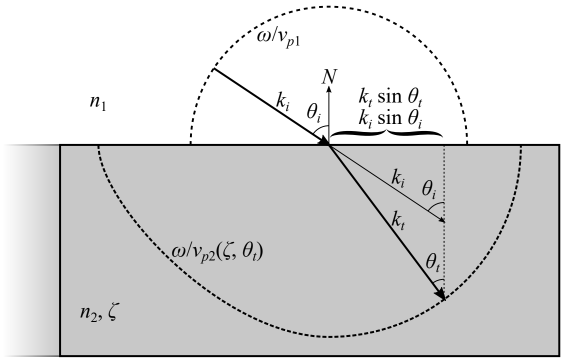


Figure 6.13 Refraction into a moving material. The source material is assumed to be at rest, thus having a circular contour for index of refraction in the upper hemisphere, while the target material is in motion, distorting the contour in the lower hemisphere.

Remember that the change in $\mathbf{k}$ is along the direction of the surface normal; therefore, the refracted wavevector, $\mathbf{k}_t$, is given by the intersection of the lower contour with the vertical line passing through $\mathbf{k}_i$.

References

[1] M. A. Morgan, *Principles of RF and Microwave Design*. Norwood, MA: Artech House, November 2019.

[2] D. M. Pozar, *Microwave Engineering*, 4th ed. New York: Wiley, 2011.

[3] C. Balanis, *Advanced Engineering Electromagnetics*. New York: Wiley, 1989.

[4] Wikipedia. (2023, December) Covariant formulation of classical electromagnetism. https://en.wikipedia.org/wiki/Covariant_formulation_of_classical_electromagnetism.

[5] M. Jannsen and J. Stachel, "Optics and electrodynamics of moving bodies," *Max Planck Institute for the History of Science*, pp. 1–37, 2004.

[6] Wikipedia. (2023, March) Fizeau experiment. https://en.wikipedia.org/wiki/Fizeau_experiment.

[7] Wikipedia. (2023, December) Michelson-Morley experiment. https://en.wikipedia.org/wiki/Michelson-Morley_experiment.

[8] Wikipedia. (2024, January) Cherenkov radiation. https://en.wikipedia.org/wiki/Cherenkov_radiation.

[9] J. M. Charap, *Covariant Electrodynamics: A Concise Guide*. Maryland: Johns Hopkins University Press, 2011.

Chapter 7

Guided Waves

In this chapter, we turn our attention to the properties of Faraday-field propagation within waveguiding structures. It is only natural that a preferred reference frame emerges from the setup of the problem, namely the rest frame of the guiding structure itself, but we will be interested in how the field and wave properties change in other reference frames, especially those resulting from longitudinal boosts, under which the guide geometry is typically invariant.

7.1 RECTANGULAR WAVEGUIDE

Let us consider for our first case the rectangular waveguide shown in Figure 7.1. We assume that its walls are perfect electric conductors and that the broad wall has interior dimension a, while the narrow wall has interior dimension b. For simplicity, we will derive our initial solutions under the assumption that the broad wall is parallel to the γ_1 direction, the narrow wall is parallel to γ_2, and γ_3 is the longitudinal direction of wave propagation. For now, we allow the possibility

Figure 7.1 Geometry of a rectangular waveguide.

that the waveguide is filled with a dielectric material, having proper constitutive parameters μ and ε that are both isotropic and homogeneous.

7.1.1 Transverse and Longitudinal Fields

Let us once again use the canonical polar form for the Faraday field, only this time with both possible rotational sign conventions included,[1]

$$\mathcal{F} = \tfrac{1}{2}\mathbf{f}e^{i\varphi} + \tfrac{1}{2}\mathbf{g}e^{-i\varphi} \tag{7.1}$$

Completely general solutions to the fields inside a waveguide are diverse, but they can always be written as the sum of the contributions from an arbitrary number of propagating modes that obey certain simplifying assumptions. For example, we assume that all variations in the γ_{30} plane (the temporal and longitudinal coordinates) are embodied in the complexion, φ. The simple bivectors, $\mathbf{f}$ and $\mathbf{g}$, must then vary in the transverse plane only

$$\mathcal{F} = \tfrac{1}{2}\mathbf{f}\left(x_{\mathbf{t}}\right)e^{i\varphi(x_{\mathbf{z}})} + \tfrac{1}{2}\mathbf{g}\left(x_{\mathbf{t}}\right)e^{-i\varphi(x_{\mathbf{z}})} \tag{7.2}$$

where the subscripts $\mathbf{t}$ and $\mathbf{z}$ in this context denote projection of the relevant vectors onto the transverse (γ_{21}) and temporal-longitudinal (γ_{30}) spacetime planes,

$$x_{\mathbf{t}} = \left(x \cdot \gamma_{21}\right)\gamma_{21}^{-1} = -\left(x \cdot \gamma_{21}\right)\gamma_{21} \tag{7.3a}$$

$$x_{\mathbf{z}} = \left(x \cdot \gamma_{30}\right)\gamma_{30}^{-1} = \left(x \cdot \gamma_{30}\right)\gamma_{30} \tag{7.3b}$$

Note that this does not mean $\mathbf{f}$ has no directed component in the γ_{30} plane, only that it does not vary along those dimensions. It is useful to subdivide the spacetime gradient along the same coordinate axes,[2]

$$\blacksquare = \blacksquare_{\mathbf{t}} + \blacksquare_{\mathbf{z}} \tag{7.4}$$

Finally, we assume that the region inside the waveguide is source-free (which does not preclude the possibility of sources, i.e., charges and currents, in the conducting

1 A superposition of these terms will be necessary to derive complete and valid mode solutions, whose field bivectors can no longer be purely transverse.

2 We are using the filled-square gradient here to allow for the possibility that the waveguide has a dielectric filler. The similarity of the macroscopic and microscopic equations in source-free media ensures that this entire derivation could be repeated for an empty waveguide simply by substituting $\blacksquare \to \square$, $\mathcal{F} \to cF$, and $n \to 1$.

walls). We may thus write Maxwell's equation as follows,

$$\blacksquare\mathcal{F} = \tfrac{1}{2}\left(\blacksquare_{\mathbf{t}}\mathbf{f} + (\blacksquare_{\mathbf{z}}\varphi)\, i\mathbf{f}\right) e^{i\varphi} + \tfrac{1}{2}\left(\blacksquare_{\mathbf{t}}\mathbf{g} - (\blacksquare_{\mathbf{z}}\varphi)\, i\mathbf{g}\right) e^{-i\varphi} = 0 \tag{7.5}$$

To satisfy this equation for all φ, we must have

$$\blacksquare_{\mathbf{t}}\mathbf{f} = -(\blacksquare_{\mathbf{z}}\varphi)i\mathbf{f} = i(\blacksquare_{\mathbf{z}}\varphi)\mathbf{f} \tag{7.6a}$$

$$\blacksquare_{\mathbf{t}}\mathbf{g} = (\blacksquare_{\mathbf{z}}\varphi)i\mathbf{g} \tag{7.6b}$$

Note that the left sides of both equations above depend only on the transverse coordinates. We therefore require that the terms on the right side involving elements with longitudinal variation must be constant,

$$\blacksquare_{\mathbf{z}}\varphi = L \tag{7.7a}$$

$$\therefore \varphi = K \cdot x \tag{7.7b}$$

where (for now we are assuming the rest frame),

$$K = n^{-1}L_U + L_{\not{U}} = n^{-1}L_{c\gamma_0} + L_{\cancel{c\gamma_0}} \tag{7.8}$$

It is evident that both K and L are vectors that lie in the γ_{30} plane,

$$L = k\gamma_0 + \beta\gamma_3 \tag{7.9a}$$

$$K = \tfrac{k}{n}\gamma_0 + \beta\gamma_3 \tag{7.9b}$$

We should not confuse β in this context with the relative velocity, or Jackson number, introduced in Chapter 2; the symbols k and β are used here for consistency with the unguided wavenumber in the material ($k = n\omega/c$) and the lossless propagation constant, respectively, which are common parameters in microwave analysis. We may also recognize that the normalization of L is simply the cutoff wavenumber,

$$L^2 = k^2 - \beta^2 = k_c^2 \tag{7.10}$$

Returning to our equation for $\mathbf{f}$, we have

$$\blacksquare_{\mathbf{t}}\mathbf{f} = iL\mathbf{f} \tag{7.11a}$$

$$\blacksquare_{\mathbf{t}}^2\mathbf{f} = iL\blacksquare_{\mathbf{t}}\mathbf{f} = L^2\mathbf{f} = k_c^2\mathbf{f} \tag{7.11b}$$

$$\therefore \left(\blacksquare_{\mathbf{t}}^2 - k_c^2\right)\mathbf{f} = 0 \tag{7.11c}$$

This is the *Helmholtz equation* [1–3] expressed in the language of spacetime algebra. We may convert it into a slightly more recognizable form by making the substitution,

$$\blacksquare_{\mathbf{t}}^2 = -\boldsymbol{\nabla}_{\mathbf{t}}^2 \tag{7.12}$$

where a minus sign is applied because the transverse plane over which differentiation occurs is spatial, and thus has a negative square norm in the chosen metric signature. The Helmholtz equation then becomes

$$\left(\boldsymbol{\nabla}_{\mathbf{t}}^2 + k_c^2\right)\mathbf{f} = 0 \tag{7.13}$$

Either way, it is a second-order differential equation comprising a scalar operator acting on each of the components of a simple bivector independently. As it is customary to categorize propagating modes in waveguides by the presence or absence of longitudinal electric and magnetic field components — or, in the language of spacetime algebra, by the longitudinally directed polar (real) and axial (imaginary) bivector parts — we will find it useful to segregate the Helmholtz equation into parts that operate on the transverse and longitudinal components separately,[3]

$$\left(\blacksquare_{\mathbf{t}}^2 - k_c^2\right)\mathbf{f}_t = 0 \tag{7.14a}$$

$$\left(\blacksquare_{\mathbf{t}}^2 - k_c^2\right)\mathbf{f}_z = 0 \tag{7.14b}$$

Applying this same decomposition to (7.11a), we have

$$\blacksquare_{\mathbf{t}}\left(\mathbf{f}_t + \mathbf{f}_z\right) = iL\left(\mathbf{f}_t + \mathbf{f}_z\right) \tag{7.15}$$

Certain pairs of terms in this equation are orthogonal to one another, allowing us to separate them. For example, $\blacksquare_{\mathbf{t}}\mathbf{f}_t$ results in only longitudinal components,

$$\blacksquare_{\mathbf{t}}\mathbf{f}_t \in \begin{Bmatrix}\gamma_1\\ \gamma_2\end{Bmatrix}\begin{Bmatrix}\sigma_1\\ \sigma_2\\ i\sigma_1\\ i\sigma_2\end{Bmatrix} = \begin{Bmatrix}\gamma_1\\ \gamma_2\end{Bmatrix}\begin{Bmatrix}\gamma_{10}\\ \gamma_{20}\\ \gamma_{32}\\ \gamma_{13}\end{Bmatrix} = \begin{Bmatrix}\gamma_0\\ \gamma_3\\ i\gamma_0\\ i\gamma_3\end{Bmatrix} \tag{7.16}$$

3 Unlike the vector projections $x_{\mathbf{t}}$, $x_{\mathbf{z}}$, $\blacksquare_{\mathbf{t}}$, and $\blacksquare_{\mathbf{z}}$, the transverse and longitudinal bivector field quantities $\mathbf{f}_t$ and $\mathbf{f}_z$ are not actually found via projection. The reader may verify that the projection of a bivector $\mathbf{f}$ onto $\gamma_{30} = \sigma_3$ has only the σ_3 component, and lacks the $i\sigma_3$ component, which is desired in this case. Even worse, projection onto γ_3 yields the longitudinal σ_3 part along with the transverse $i\sigma_1$ and $i\sigma_2$ parts. Nevertheless, we shall simply use the notation $\mathbf{f}_z$ to represent the σ_3 and $i\sigma_3$ components of $\mathbf{f}$, and $\mathbf{f}_t$ to represent the σ_1, σ_2, $i\sigma_1$, and $i\sigma_2$ components, despite the absence of a compact mathematical formula for extracting them.

Similarly, $iL\mathbf{f}_z$ also yields longitudinal components exclusively. Retaining only those terms that produce transverse components, we are left with

$$\blacksquare_{\mathbf{t}}\mathbf{f}_z = iL\mathbf{f}_t \tag{7.17}$$

which can be inverted to solve for $\mathbf{f}_t$,

$$\mathbf{f}_t = \tfrac{iL}{k_c^2}\blacksquare_{\mathbf{t}}\mathbf{f}_z \tag{7.18}$$

Specific solutions for $\mathbf{f}$ may therefore be derived by first solving (7.14b), the Helmholtz equation for the longitudinal components, and then using those solutions in (7.18) to find the associated transverse fields. A similar procedure yields the following expression for the transverse components of $\mathbf{g}$ in terms of its longitudinal field (which must also obey the Helmholtz equation),

$$\mathbf{g}_t = -\tfrac{iL}{k_c^2}\blacksquare_{\mathbf{t}}\mathbf{g}_z \tag{7.19}$$

Both the Helmholtz equation and the boundary conditions for a closed metallic waveguide can be satisfied for a single mode if the total longitudinal field is either purely polar or purely axial. That, in turn, requires that the longitudinal components of $\mathbf{f}$ and $\mathbf{g}$ are identical,

$$\mathbf{g}_z = \mathbf{f}_z \tag{7.20}$$

Otherwise, the complex phase rotation in (7.1) would convert polar fields to axial fields and axial to polar as the wave propagated. We thus find that $\mathbf{g}_t = -\mathbf{f}_t$ and the complete Faraday field may be written as,

$$\mathcal{F} = \mathcal{F}_t + \mathcal{F}_z = \mathbf{f}_z\cos\varphi + i\mathbf{f}_t\sin\varphi = \left(\cos\varphi - \tfrac{1}{k_c^2}L\blacksquare_{\mathbf{t}}\sin\varphi\right)\mathbf{f}_z \tag{7.21}$$

It will pay dividends later on to make note of the following expansion,

$$-L\blacksquare_{\mathbf{t}} = -\left(k\gamma_0 + \beta\gamma_3\right)\left(\gamma^1\partial_1 + \gamma^2\partial_2\right) \tag{7.22a}$$

$$= -\left(k\gamma_0 + \beta\gamma_3\right)\gamma^0\gamma_0\left(\gamma^1\partial_1 + \gamma^2\partial_2\right) = -\left(k + \beta\sigma_3\right)\left(\sigma_1\partial_1 + \sigma_2\partial_2\right) \tag{7.22b}$$

$$= \left(\beta + k\sigma_3\right)\left(\sigma_1\partial_1 + \sigma_2\partial_2\right)\sigma_3^{-1} = \left(\beta + k\sigma_3\right)\boldsymbol{\nabla}_t\sigma_3^{-1} \tag{7.22c}$$

By using the del operator, $\boldsymbol{\nabla}_t$, we have put this into a form that is independent of the coordinate system. In Cartesian coordinates, it can be expanded as

$$-L\blacksquare_{\mathbf{t}} = \beta\left(\sigma_1\partial_1 + \sigma_2\partial_2\right)\sigma_3^{-1} + ik\left(\sigma_2\partial_1 - \sigma_1\partial_2\right)\sigma_3^{-1} \tag{7.23}$$

Table 7.1
Equations for Arbitrary Uniform Closed Metallic Waveguides

Parameter	Symbol	Expression
Unguided wavenumber	k	$= \frac{n\omega}{c} = \omega\sqrt{\mu\varepsilon}$
Four-wavevector*	K	$= \frac{k}{n}\gamma_0 + \beta\gamma_3$
Phase gradient*	L	$= k\gamma_0 + \beta\gamma_3$
Invariant phase	φ	$= K \cdot x$
Lossless propagation constant	β	$= \sqrt{k^2 - k_c^2}$
Faraday field	$\mathcal{F}$	$= \left(\cos\varphi - \frac{1}{k_c^2} L\blacksquare_{\mathbf{t}} \sin\varphi\right) \mathbf{f}_z$

*For the worldline defined by γ_0 assuming γ_3 is the direction of propagation.

Expansions in other coordinate systems (e.g., cylindrical or spherical) can be derived with the aid of the scalar gradient formulas from three-dimensional vector calculus found in Appendix B.

The results of this section are summarized in Table 7.1. It is worth pointing out that no part of this derivation utilized the specific geometry of the rectangular waveguide in Figure 7.1. We have assumed only that it has a longitudinal direction, that it is uniform in that direction, and that it is homogeneously filled with dielectric (or air). We will employ these same results again in Section 7.2 when discussing a waveguide having an entirely different cross-sectional geometry, namely circular.

7.1.2 Transverse Axial Modes

Traditionally, modes for which $\mathcal{F}_z$ is polar (real) were called *E-modes* or *transverse magnetic* (TM) modes in recognition of the usual interpretation that the polar bivector part of the Faraday field represented the electric field, and thus only the axial part (the magnetic field) was entirely transverse. In my view, this runs counter to the spirit of the spacetime interpretation of the Faraday field as a unified, frame-independent quantity, so I will refer to them instead as *transverse axial* (TA) modes, since their axial components lie exclusively in the transverse spatial plane. It can be acknowledged that this distinction is merely semantic, but to my way of thinking, this nomenclature does not so blatantly discard the unification of electric and magnetic fields, which is paramount to this work.

If $\mathcal{F}_z$ is to be entirely polar, then $\mathbf{f}_z$ may only have a σ_3 component,

$$\mathbf{f}_z = f_z \sigma_3 \tag{7.24}$$

The Helmholtz equation for the longitudinal field (7.14b) then reduces to a scalar form which can be solved using traditional methods,

$$\blacksquare_{\mathbf{t}}^2 f_z = k_c^2 f_z \tag{7.25}$$

Its solutions take on the form of sinusoidal functions along the differentiated coordinates,

$$f_z = k_c^2 F_0 \sin\left(K_1 \cdot x\right) \sin\left(K_2 \cdot x\right) \tag{7.26}$$

where

$$K_1 = k_1 \gamma^1 \tag{7.27a}$$

$$K_2 = k_2 \gamma^2 \tag{7.27b}$$

$$k_1^2 + k_2^2 = k_c^2 \tag{7.27c}$$

and the prefactor ($k_c^2 F_0$) has been selected to simplify the final expressions. The boundary condition for perfect conductors on the narrow walls of the rectangular waveguide require that $f_z = 0$ at $x \cdot \gamma^1 = 0$ and $x \cdot \gamma^1 = a$, since

$$\mathbf{f}_z \wedge N = \pm \left(\sigma_3 \wedge \gamma_1\right) f_z = \pm \left(i \gamma_2\right) f_z = 0 \tag{7.28}$$

We may thus conclude that

$$k_1 = \frac{m\pi}{a} \tag{7.29}$$

for some positive integer m. Similarly, the boundary condition on the broad walls at $x \cdot \gamma^2 = 0$ and $x \cdot \gamma^2 = b$ require that

$$k_2 = \frac{n\pi}{b} \tag{7.30}$$

for some integer n (not to be confused with the index of refraction of the interior dielectric). Note that both m and n must be greater than or equal to 1; otherwise, the Faraday field would be zero.

We now have enough information to write down the complete Faraday field of a transverse axial mode with indices m and n, which shall be labeled the TA_{mn} mode,

$$\mathcal{F} = \left(\cos\varphi - \tfrac{1}{k_c^2} L \blacksquare_{\mathbf{t}} \sin\varphi\right) \mathbf{f}_z \tag{7.31a}$$

$$= F_0 \left(k_c^2 \cos\varphi - L\blacksquare_{\mathbf{t}} \sin\varphi\right) \sigma_3 \sin\left(K_1 \cdot x\right) \sin\left(K_2 \cdot x\right) \tag{7.31b}$$

$$= F_0 \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ i\sigma_1 \\ i\sigma_2 \\ i\sigma_3 \end{pmatrix}^T \begin{pmatrix} k_1\beta \cos\left(K_1 \cdot x\right) \sin\left(K_2 \cdot x\right) \sin\varphi \\ k_2\beta \sin\left(K_1 \cdot x\right) \cos\left(K_2 \cdot x\right) \sin\varphi \\ k_c^2 \sin\left(K_1 \cdot x\right) \sin\left(K_2 \cdot x\right) \cos\varphi \\ -k_2 k \sin\left(K_1 \cdot x\right) \cos\left(K_2 \cdot x\right) \sin\varphi \\ k_1 k \cos\left(K_1 \cdot x\right) \sin\left(K_2 \cdot x\right) \sin\varphi \\ 0 \end{pmatrix} \tag{7.31c}$$

7.1.3 Transverse Polar Modes

Traditionally, modes for which $\mathcal{F}_z$ is axial (imaginary) were called *H-modes* or *transverse electric* (TE) modes, as the axial part is associated with the magnetic field, and thus only the polar part (the electric field) was exclusively transverse. I will refer to these in this work as *transverse polar* (TP) modes. We thus write the simple bivector $\mathbf{f}_z$ as

$$\mathbf{f}_z = f_z i\sigma_3 \tag{7.32}$$

The scalar form of the Helmholtz equation is unchanged, and the solution once again lies in the sinusoidal functions. However, the boundary condition in this case is not found in the value of the field at the conducting walls, but in its slope as described in (6.121). For the narrow walls at $x \cdot \gamma^1 = 0$ and $x \cdot \gamma^1 = a$,

$$\blacksquare_N \left(f_z i\sigma_3\right)_N = \pm\left(\gamma^1 \partial_1\right)\left(f_z i\sigma_3\right) = \pm\gamma_2 \partial_1 f_z = 0 \tag{7.33a}$$

$$\therefore\ \partial_1 f_z\big|_{x\cdot\gamma^1 = 0,a} = 0 \tag{7.33b}$$

and similarly for the broad walls,

$$\partial_2 f_z\big|_{x\cdot\gamma^2 = 0,b} = 0 \tag{7.34}$$

We therefore use the cosine solution instead of the sine,

$$f_z = k_c^2 F_0 \cos\left(K_1 \cdot x\right) \cos\left(K_2 \cdot x\right) \tag{7.35}$$

where once again

$$k_1^2 + k_2^2 = k_c^2 \tag{7.36a}$$

$$k_1 = \frac{m\pi}{a} \tag{7.36b}$$

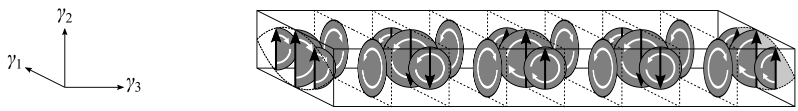


Figure 7.2 The dominant TP_{10} mode for a rectangular waveguide. The polar components are shown with the black arrows, while the axial components are shown with the gray planar patches and white arrows.

$$k_2 = \frac{n\pi}{b} \tag{7.36c}$$

The complete TP_{mn} mode may thus be written,

$$\mathcal{F} = \left(\cos\varphi - \tfrac{1}{k_c^2} L\blacksquare_{\mathbf{t}} \sin\varphi\right) \mathbf{f}_z \tag{7.37a}$$

$$= F_0 \left(k_c^2 \cos\varphi - L\blacksquare_{\mathbf{t}} \sin\varphi\right) i\sigma_3 \cos\left(K_1 \cdot x\right) \cos\left(K_2 \cdot x\right) \tag{7.37b}$$

$$= F_0 \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ i\sigma_1 \\ i\sigma_2 \\ i\sigma_3 \end{pmatrix}^T \begin{pmatrix} -k_2 k \cos\left(K_1 \cdot x\right) \sin\left(K_2 \cdot x\right) \sin\varphi \\ k_1 k \sin\left(K_1 \cdot x\right) \cos\left(K_2 \cdot x\right) \sin\varphi \\ 0 \\ -k_1 \beta \sin\left(K_1 \cdot x\right) \cos\left(K_2 \cdot x\right) \sin\varphi \\ -k_2 \beta \cos\left(K_1 \cdot x\right) \sin\left(K_2 \cdot x\right) \sin\varphi \\ k_c^2 \cos\left(K_1 \cdot x\right) \cos\left(K_2 \cdot x\right) \cos\varphi \end{pmatrix} \tag{7.37c}$$

The bivector field for the TP_{10} mode is illustrated in Figure 7.2.[4] The mathematical results derived here for all modes in rectangular waveguide are summarized in Table 7.2.

In this case, either m or n may be zero, but not both. Assuming that a is the larger dimension, this means that the smallest critical wavenumber k_c for any mode is that associated with mode TP_{10},

$$k_c = \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} = \frac{\pi}{a} \quad \text{for} \quad (m, n) = (1, 0) \tag{7.38}$$

4 The reader may notice that this figure is a little different than what is commonly drawn in most microwave textbooks. That is because, traditionally, both the electric and magnetic fields are considered vectors, or directed line segments drawn with arrows. In the context of special relativity, however, we now understand that the Faraday field is a bivector, or directed plane segment, from which the apparent electric and magnetic vectors are derived. We therefore draw the spatial components of the Faraday field as a two-dimensional surface (the gray patches in Figure 7.2) with orientation indicated by the circulating arrows.

Table 7.2
Equations for Rectangular Waveguide

Parameter	Symbol	Expression
Horizontal wavenumber	k_1	$= \frac{m\pi}{a}$
Vertical wavenumber	k_2	$= \frac{n\pi}{b}$
Cutoff wavenumber	k_c	$= \sqrt{k_1^2 + k_2^2}$
Dominant mode ($a > b$)	TP_{10}	
TA field	$\begin{pmatrix} \mathcal{F}_{\sigma_1} \\ \mathcal{F}_{\sigma_2} \\ \mathcal{F}_{\sigma_3} \\ \mathcal{F}_{i\sigma_1} \\ \mathcal{F}_{i\sigma_2} \\ \mathcal{F}_{i\sigma_3} \end{pmatrix}$	$= F_0 \begin{pmatrix} k_1\beta \cos(K_1 \cdot x) \sin(K_2 \cdot x) \sin\varphi \\ k_2\beta \sin(K_1 \cdot x) \cos(K_2 \cdot x) \sin\varphi \\ k_c^2 \sin(K_1 \cdot x) \sin(K_2 \cdot x) \cos\varphi \\ -k_2 k \sin(K_1 \cdot x) \cos(K_2 \cdot x) \sin\varphi \\ k_1 k \cos(K_1 \cdot x) \sin(K_2 \cdot x) \sin\varphi \\ 0 \end{pmatrix}$
TP field	$\begin{pmatrix} \mathcal{F}_{\sigma_1} \\ \mathcal{F}_{\sigma_2} \\ \mathcal{F}_{\sigma_3} \\ \mathcal{F}_{i\sigma_1} \\ \mathcal{F}_{i\sigma_2} \\ \mathcal{F}_{i\sigma_3} \end{pmatrix}$	$= F_0 \begin{pmatrix} -k_2 k \cos(K_1 \cdot x) \sin(K_2 \cdot x) \sin\varphi \\ k_1 k \sin(K_1 \cdot x) \cos(K_2 \cdot x) \sin\varphi \\ 0 \\ -k_1\beta \sin(K_1 \cdot x) \cos(K_2 \cdot x) \sin\varphi \\ -k_2\beta \cos(K_1 \cdot x) \sin(K_2 \cdot x) \sin\varphi \\ k_c^2 \cos(K_1 \cdot x) \cos(K_2 \cdot x) \cos\varphi \end{pmatrix}$

This can be associated with the *cutoff frequency*,

$$\omega_c = \frac{k_c}{\sqrt{\mu\varepsilon}} = \frac{k_c c}{n} = \frac{\pi c}{an} \tag{7.39}$$

Below this frequency, no mode can propagate; the waveguide is said to be cutoff. Between this frequency and the cutoff for the next higher-order mode, the waveguide is *single-moded*. This is the range most often associated with the usable frequencies of the waveguide (at best, manufacturing tolerances and losses limit practical waveguide operation to an even narrower range of frequencies). Most standard waveguides have a 2:1 aspect ratio ($a = 2b$), affording them an octave of single-mode operation, the most that can be achieved with any aspect ratio. The next higher modes, TP_{20} and TP_{01}, have the same cutoff frequency in 2:1 waveguide, and are thus said to be *degenerate* [1].

Figure 7.3 Geometry of a circular waveguide and the system of cylindrical coordinates that will be used to describe it.

7.2 CIRCULAR WAVEGUIDE

Up to now, everything we have derived about the behavior of the Faraday field and special relativity has implicitly assumed an underlying Cartesian coordinate system. While fundamentally valid, engineers know that other systems of coordinates are immensely advantageous for certain geometries having nonplanar boundaries.

7.2.1 Cylindrical Coordinates

The present case is one example: a circular waveguide such as that shown in Figure 7.3. This structure is much better suited to cylindrical coordinates. In the context of four-dimensional spacetime, the temporal axis, γ_0, is unchanged, as is the longitudinal axis, γ_3, but we shall replace the transverse coordinates γ_1 and γ_2 with the radial and angular coordinates ρ_1 and ϕ_2, respectively, as shown. The associated bivector planes resulting from extending these axes through time shall be denoted ρ_{10} and ϕ_{20}. Note that, as drawn, the new coordinates are numbered so as to ensure a right-handed coordinate system,

$$\rho_{10} \times \phi_{20} = \gamma_{30} \tag{7.40}$$

It can further be shown that

$$\gamma_3 \wedge \phi_2 = i\rho_{10} \tag{7.41a}$$

$$\rho_1 \wedge \gamma_3 = i\phi_{20} \tag{7.41b}$$

$$\phi_2 \wedge \rho_1 = i\sigma_3 \tag{7.41c}$$

Before attempting to describe the modes in this waveguide, it will be useful to convert the expansion of $-L\blacksquare_{\mathbf{t}}$ from (7.22c) into cylindrical coordinates using the formula given in (D.65b),[5]

$$-L\blacksquare_{\mathbf{t}} = \beta\left(\rho_{10}\partial_\rho + \rho^{-1}\phi_{20}\partial_\phi\right)\sigma_3^{-1} + k\sigma_3\left(\rho_{10}\partial_\rho + \rho^{-1}\phi_{20}\partial_\phi\right)\sigma_3^{-1} \quad (7.42)$$

The factors in the second term may be simplified,

$$\sigma_3\rho_{10} = \gamma_3\gamma_0\rho_1\gamma_0 = \rho_1\gamma_3 = i\phi_{20} \quad (7.43a)$$

$$\sigma_3\phi_{20} = \gamma_3\gamma_0\phi_2\gamma_0 = -\gamma_3\phi_2 = -i\rho_{10} \quad (7.43b)$$

Therefore,

$$-L\blacksquare_{\mathbf{t}} = \beta\left(\rho_{10}\partial_\rho + \rho^{-1}\phi_{20}\partial_\phi\right)\sigma_3^{-1} + ik\left(\phi_{20}\partial_\rho - \rho^{-1}\rho_{10}\partial_\phi\right)\sigma_3^{-1} \quad (7.44)$$

7.2.2 Transverse Axial Modes

A transverse axial mode will have a polar longitudinal component $\mathbf{f}_z = f_z\sigma_3$ that obeys the scalar Helmholtz equation,

$$\left(\blacksquare_{\mathbf{t}}^2 - k_c^2\right) f_z = 0 \quad (7.45)$$

Let us translate this into the cylindrical form using the Laplacian given in (D.66),

$$\left(\boldsymbol{\nabla}_t^2 + k_c^2\right) f_z = 0 \quad (7.46a)$$

$$\left\{\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2}{\partial\phi^2} + k_c^2\right\} f_z = 0 \quad (7.46b)$$

We can solve this equation using a *separation of variables* technique [4–6]. First, we assume that its solution has the form of the product of two functions, one which depends on ρ only, and another which depends on ϕ only,

$$f_z(\rho,\phi) = R(\rho)P(\phi) \quad (7.47)$$

5 We are taking advantage of the fact that the appended factor σ_3^{-1} here will render any longitudinal field solution into a scalar or pseudoscalar, to which the gradient form of the del operator (as opposed to the divergence or curl) can then be applied.

Substituting these into (7.46b), we have

$$\left\{\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial}{\partial\rho}\right)+\frac{1}{\rho^2}\frac{\partial^2}{\partial\phi^2}+k_c^2\right\}R(\rho)P(\phi)=0 \tag{7.48a}$$

$$P\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial R}{\partial\rho}\right)+R\frac{1}{\rho^2}\frac{\partial^2 P}{\partial\phi^2}+k_c^2RP=0 \tag{7.48b}$$

$$\frac{\rho}{R}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial R}{\partial\rho}\right)+k_c^2\rho^2=-\frac{1}{P}\frac{\partial^2 P}{\partial\phi^2} \tag{7.48c}$$

The left side of this equation depends on ρ only, while the right side depends on ϕ only. Both must therefore be equal to the same constant, λ, and can be written as separate equations,

$$\frac{\rho}{R}\frac{d}{d\rho}\left(\rho\frac{dR}{d\rho}\right)+k_c^2\rho^2=\lambda \tag{7.49a}$$

$$-\frac{1}{P}\frac{d^2P}{d\phi^2}=\lambda \tag{7.49b}$$

As each equation now depends on only one variable, the partial derivatives have been replaced with full derivatives. The original differential equation is thus said to be *separable*, and λ is the *separation constant*. Let us solve the equation for $P(\phi)$ first. A slight rearrangement reveals that this is once again a manifestation of the Helmholtz equation in a single variable,

$$\frac{d^2P}{d\phi^2}+\lambda P=0 \tag{7.50}$$

The solution can be written as the sum of sines and cosines,

$$P(\phi)=P_1\sin\left(\sqrt{\lambda}\phi\right)+P_2\cos\left(\sqrt{\lambda}\phi\right) \tag{7.51}$$

or, more simply, as a single sinusoid with an arbitrary phase offset,

$$P(\phi)=P_0\cos\left(\sqrt{\lambda}(\phi+\psi)\right) \tag{7.52}$$

Because ϕ represents an azimuthal angle in space, any real solution must be periodic in that variable, so we constrain $\sqrt{\lambda}$ to be an integer, n,

$$P(\phi)=P_0\cos\left(n(\phi+\psi)\right) \tag{7.53}$$

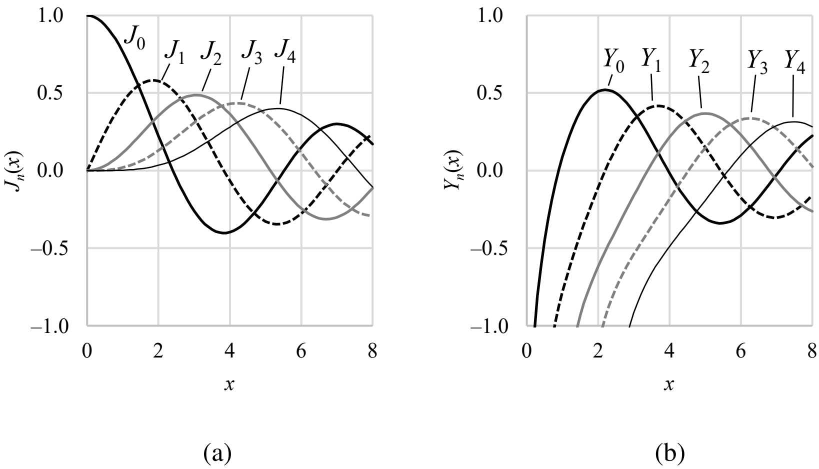


Figure 7.4 Bessel functions of (a) the first kind and (b) the second kind for orders $n = 0 \ldots 4$.

If we now use this separation constant in the equation for $R(\rho)$, we have

$$\rho^2 \frac{d^2 R}{d\rho^2} + \rho \frac{dR}{d\rho} + \left(k_c^2 \rho^2 - n^2\right) R = 0 \tag{7.54}$$

This is a well-studied differential equation known as *Bessel's equation*, with argument $x = k_c\rho$ and order n, whose solutions are the Bessel functions of the first and second kinds [1, 7],

$$R(\rho) = R_1 J_n\left(k_c\rho\right) + R_2 Y_n\left(k_c\rho\right) \tag{7.55}$$

These functions are shown in Figure 7.4. The Bessel functions of the second kind, Y_n, are infinite at $k_c\rho = 0$ and are thus not a realistic solution for this problem. We therefore use the functions of the first kind exclusively.

Putting these results together, we have for the longitudinal field component,

$$f_z = k_c^2 F_0 J_n\left(k_c\rho\right) \cos\left(n(\phi + \psi)\right) \tag{7.56}$$

The boundary conditions require that this component reaches zero at the walls of the waveguide ($\rho = a$). We may therefore determine the cutoff wavenumber as

$$k_c = \frac{z_{nm}}{a} \tag{7.57}$$

Table 7.3
Zeros (z_{nm}) and Stationary Points (e_{nm}) of the Bessel Function of the First Kind (J_n)

n	z_{n1}	z_{n2}	z_{n3}	z_{n4}	e_{n1}	e_{n2}	e_{n3}	e_{n4}
0	2.4048	5.5201	8.6537	11.7915	3.8317	7.0156	10.1735	13.3237
1	3.8317	7.0156	10.1735	13.3237	1.8412	5.3314	8.5363	11.7060
2	5.1356	8.4172	11.6198	14.7960	3.0542	6.7061	9.9695	13.1704
3	6.3802	9.7610	13.0152	16.2235	4.2012	8.0512	11.3459	14.5858
4	7.5883	11.0647	14.3725	17.6160	5.3175	9.2824	12.6819	15.9641
5	8.7715	12.3386	15.7002	18.9801	6.4156	10.5199	13.9872	17.3128

where z_{nm} is the mth zero, or root, of the nth-order Bessel function. There are infinite series for calculating the roots of J_n, the results of which are tabulated in Table 7.3 for convenience. The azimuthal harmonic is allowed to be zero ($n \geq 0$), but by convention the first zero of the Bessel function is numbered $m = 1$. Thus, the lowest-order transverse axial mode is TA_{01} with a cutoff wavenumber of $2.4048/a$.

The complete Faraday field of a transverse axial mode in circular waveguide is found using (7.21) with the differential operator expansion (7.44),

$$\mathcal{F} = \left(\cos\varphi - \tfrac{1}{k_c^2} L\blacksquare_{\mathbf{t}} \sin\varphi\right) \mathbf{f}_z \tag{7.58a}$$

$$= \left(k_c^2 \cos\varphi - L\blacksquare_{\mathbf{t}} \sin\varphi\right) \sigma_3 F_0 J_n\left(k_c\rho\right) \cos\left(n(\phi+\psi)\right) \tag{7.58b}$$

$$= F_0 \begin{pmatrix} \rho_{10} \\ \phi_{20} \\ \sigma_3 \\ i\rho_{10} \\ i\phi_{20} \\ i\sigma_3 \end{pmatrix}^T \begin{pmatrix} \beta k_c J_n'\left(k_c\rho\right) \cos\left(n(\phi+\psi)\right) \sin\varphi \\ -n\beta\rho^{-1} J_n\left(k_c\rho\right) \sin\left(n(\phi+\psi)\right) \sin\varphi \\ k_c^2 J_n\left(k_c\rho\right) \cos\left(n(\phi+\psi)\right) \cos\varphi \\ nk\rho^{-1} J_n\left(k_c\rho\right) \sin\left(n(\phi+\psi)\right) \sin\varphi \\ kk_c J_n'\left(k_c\rho\right) \cos\left(n(\phi+\psi)\right) \sin\varphi \\ 0 \end{pmatrix} \tag{7.58c}$$

7.2.3 Transverse Polar Modes

The scalar Helmholtz equation for transverse polar modes is identical to that for the transverse axial modes described in Section 7.2.2, and the solution is the same,

$$f_z = k_c^2 F_0 J_n\left(k_c\rho\right) \cos\left(n(\phi+\psi)\right) \tag{7.59}$$

However, in this case it is the derivative of the longitudinal field that is constrained at the boundary, not its value. The cutoff wavenumber is therefore determined not by the roots of the Bessel function, but by its stationary points, or *extrema*, where the slope becomes zero,

$$k_c = \frac{e_{nm}}{a} \tag{7.60}$$

The stationary points of the Bessel function are also tabulated in Table 7.3. Note that, somewhat counterintuitively, the TP_{11} mode has a lower cutoff wavenumber $(1.8412/a)$ than TP_{01} $(3.8317/a)$,[6] and is also lower than that for TA_{01}, making TP_{11} the dominant mode of the waveguide.

The complete field solutions are then found in the usual way,

$$\mathcal{F} = \left(\cos\varphi - \tfrac{1}{k_c^2} L\blacksquare_{\mathbf{t}} \sin\varphi\right) \mathbf{f}_z \tag{7.61a}$$

$$= \left(k_c^2 \cos\varphi - L\blacksquare_{\mathbf{t}} \sin\varphi\right) i\sigma_3 F_0 J_n\left(k_c\rho\right) \cos\left(n(\phi+\psi)\right) \tag{7.61b}$$

$$= F_0 \begin{pmatrix} \rho_{10} \\ \phi_{20} \\ \sigma_3 \\ i\rho_{10} \\ i\phi_{20} \\ i\sigma_3 \end{pmatrix}^T \begin{pmatrix} -nk\rho^{-1} J_n\left(k_c\rho\right) \sin\left(n(\phi+\psi)\right) \sin\varphi \\ -kk_c J_n'\left(k_c\rho\right) \cos\left(n(\phi+\psi)\right) \sin\varphi \\ 0 \\ \beta k_c J_n'\left(k_c\rho\right) \cos\left(n(\phi+\psi)\right) \sin\varphi \\ -n\beta\rho^{-1} J_n\left(k_c\rho\right) \sin\left(n(\phi+\psi)\right) \sin\varphi \\ k_c^2 J_n\left(k_c\rho\right) \cos\left(n(\phi+\psi)\right) \cos\varphi \end{pmatrix} \tag{7.61c}$$

These mode solutions for circular waveguide are listed in Table 7.4. A comparison of the dominant mode solutions for rectangular and circular waveguide are shown in Figure 7.5.

7.3 DISPERSION

The foregoing mode solutions were implicitly derived in the rest frame of the guiding structures, and the results should at least be familiar to those readers who have studied waveguides using classical techniques. A more novel aspect which can be explored in the context of spacetime is how the wave propagation properties vary in reference frames where the guiding structure is in motion. Waveguides inevitably exhibit some interesting properties in this area as the phase velocity may exceed the speed of light, even in the rest frame, at some frequencies for dielectric-filled waveguides, and at all frequencies for empty waveguides.

6 The true first stationary point at $x = 0$ for the zeroth-order Bessel function does not lead to a nontrivial mode solution, and is thus by convention omitted from the table.

Table 7.4
Equations for Circular Waveguide

Parameter	Symbol	Expression
Cutoff wavenumber (TA)	k_c	$= \frac{z_{nm}}{a}$
Cutoff wavenumber (TP)	k_c	$= \frac{e_{nm}}{a}$
Dominant mode	TP_{11}	
TA mode field	$\begin{pmatrix} \mathcal{F}_{\rho 10} \\ \mathcal{F}_{\phi 20} \\ \mathcal{F}_{\sigma 3} \\ \mathcal{F}_{i\rho 10} \\ \mathcal{F}_{i\phi 20} \\ \mathcal{F}_{i\sigma 3} \end{pmatrix}$	$= F_0 \begin{pmatrix} \beta k_c J_n' (k_c \rho) \cos(n(\phi+\psi)) \sin\varphi \\ -n\beta\rho^{-1} J_n (k_c\rho) \sin(n(\phi+\psi)) \sin\varphi \\ k_c^2 J_n (k_c\rho) \cos(n(\phi+\psi)) \cos\varphi \\ nk\rho^{-1} J_n (k_c\rho) \sin(n(\phi+\psi)) \sin\varphi \\ kk_c J_n' (k_c\rho) \cos(n(\phi+\psi)) \sin\varphi \\ 0 \end{pmatrix}$
TP mode field	$\begin{pmatrix} \mathcal{F}_{\rho 10} \\ \mathcal{F}_{\phi 20} \\ \mathcal{F}_{\sigma 3} \\ \mathcal{F}_{i\rho 10} \\ \mathcal{F}_{i\phi 20} \\ \mathcal{F}_{i\sigma 3} \end{pmatrix}$	$= F_0 \begin{pmatrix} -nk\rho^{-1} J_n (k_c\rho) \sin(n(\phi+\psi)) \sin\varphi \\ -kk_c J_n' (k_c\rho) \cos(n(\phi+\psi)) \sin\varphi \\ 0 \\ \beta k_c J_n' (k_c\rho) \cos(n(\phi+\psi)) \sin\varphi \\ -n\beta\rho^{-1} J_n (k_c\rho) \sin(n(\phi+\psi)) \sin\varphi \\ k_c^2 J_n (k_c\rho) \cos(n(\phi+\psi)) \cos\varphi \end{pmatrix}$

7.3.1 Unfilled Waveguides

Regardless of the cross-section of a smooth-walled waveguide (rectangular, circular, or otherwise) the propagation constant derives from the norm of the phase gradient, L,

$$\beta = \sqrt{k^2 - k_c^2} \tag{7.62}$$

where the unguided wavenumber is $k = n\omega/c$ for proper index of refraction, n. There is thus a hyperbolic relationship between the propagation constant and the frequency.

Transformation of the wavevector, K, for longitudinal boosts is also hyperbolic, as previewed in Section 3.4,

$$K = \tfrac{\omega}{c}\gamma_0 + \beta\gamma_3 \tag{7.63a}$$

$$K' = \left(\tfrac{\omega}{c}\cosh\zeta - \beta\sinh\zeta\right)\gamma_0 + \left(\beta\cosh\zeta - \tfrac{\omega}{c}\sin\zeta\right)\gamma_3 \tag{7.63b}$$

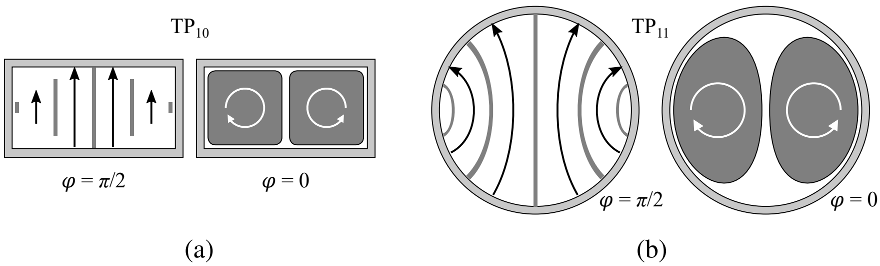


Figure 7.5 Comparison of the dominant modes in (a) rectangular waveguide and (b) circular waveguide, at two different phases, φ. The waveguides have been scaled so that they would each have the same cutoff frequency. The solid black arrows represent polar bivectors, while the axial bivector planes, seen on edge at $\varphi = \pi/2$, are in gray with white arrows.

$$\tfrac{\omega'}{c} = \tfrac{\omega}{c}\cosh\zeta - \beta\sinh\zeta \tag{7.63c}$$

$$\beta' = \beta\cosh\zeta - \tfrac{\omega}{c}\sinh\zeta = \sqrt{\left(\beta\cosh\zeta - \tfrac{\omega}{c}\sinh\zeta\right)^2} \tag{7.63d}$$

$$= \sqrt{\beta^2\cosh^2\zeta - 2\beta\tfrac{\omega}{c}\sinh\zeta\cosh\zeta + \left(\tfrac{\omega}{c}\right)^2\sinh^2\zeta} \tag{7.63e}$$

$$= \sqrt{\left(\tfrac{\omega}{c}\right)^2\cosh^2\zeta - 2\beta\tfrac{\omega}{c}\sinh\zeta\cosh\zeta + \beta^2\sinh^2\zeta - \left(\left(\tfrac{\omega}{c}\right)^2 - \beta^2\right)} \tag{7.63f}$$

$$= \sqrt{\left(\tfrac{\omega'}{c}\right)^2 - \left(\left(\tfrac{\omega}{c}\right)^2 - \beta^2\right)} = \sqrt{\left(\tfrac{k'}{n}\right)^2 - \left(\left(\tfrac{k}{n}\right)^2 - \beta^2\right)} \tag{7.63g}$$

If the waveguide is unfilled ($n = 1$) then transformed hyperbolic dispersion curve is congruent to the original,

$$\beta' = \sqrt{(k')^2 - k_c^2} \tag{7.64}$$

This is illustrated on a dispersion diagram in Figure 7.6(a). Modes that were near cutoff in the rest frame have passed through the frequency axis after the boost and are now traveling backward. This is further illustrated by the phase plotted on the Minkowski diagram of Figure 7.6(b). Whereas the x axis originally cut backward through the phase history of the wave, it now cuts forward, such that wavefronts appear to be moving toward the observer as time progresses (picture in your mind the x' axis sliding vertically through the plot, effectively moving forward in time, while its intercepts with the wave peaks track backward toward ct').

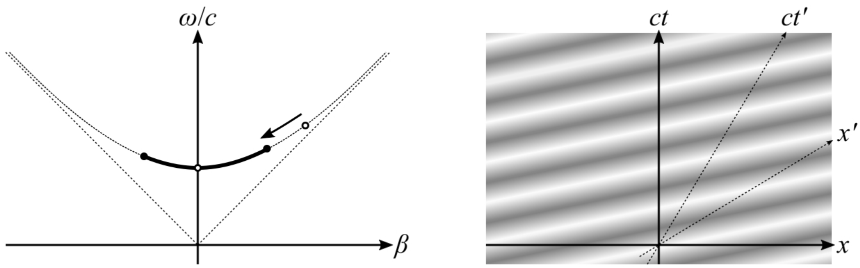


Figure 7.6 A longitudinal boost applied to the modes of an unfilled waveguide. (a) Dispersion diagram, showing a selected frequency band from cutoff to a higher frequency shifting leftward along its own hyperbolic trajectory. Modes that were previously at or near cutoff now appear as waves propagating backward. (b) Lorentz phase on a Minkowski diagram for one such wave that was originally just above cutoff (i.e., the x axis cuts backward through the phase fringes). After the boost, the x' axis cuts forward through phase fringes, indicating a backward traveling wave.

The congruence of the original and transformed hyperbolic dispersion curves is to be expected since the waves exist in vacuum and the geometry of the waveguide (assumed to be infinitely long) is unchanged by Lorentz contraction. Individual waves themselves may have transformed, but the waveguide itself effectively has not.

7.3.2 Dielectric-Filled Waveguides

The situation is very much different in a dielectric-filled waveguide. As shown in the dispersion diagram of Figure 7.7(a), the hyperbolic dispersion curve in the rest frame is no longer tangent to the diagonal asymptote for lightlike waves, but instead is tangent to a shallower asymptote for unguided waves in the medium traveling at speed $v_p = 1/\sqrt{\mu\varepsilon}$. There is a point, then, at which the guided waves transition from the usual fast waves to slow waves. At that point, they are traveling at exactly speed c, marked by its intersection with the lightlike diagonal.

The boost transformation (indicated by the gray shading) once again produces a hyperbolic dispersion curve, but it is no longer congruent to the original. Slow forward-traveling waves are retarded (the observer is catching up to the wavefronts). Fast forward-traveling waves paradoxically move even faster as they approach a kind of inverse cutoff where the phase velocity is infinite. The intercept with the ω/c axis is no longer horizontal. Fast backward-traveling waves decelerate while slow

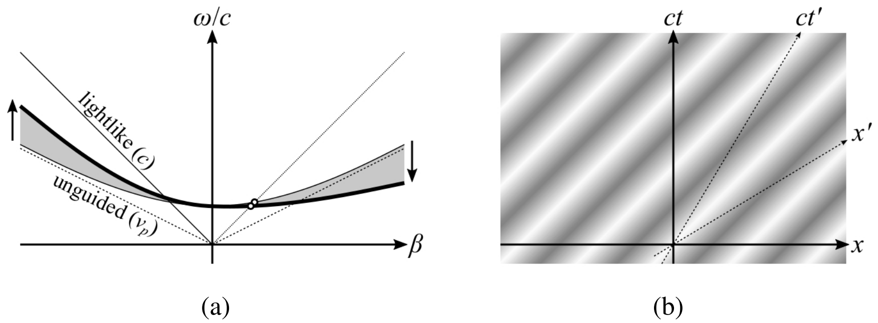


Figure 7.7 A longitudinal boost applied to the modes of a dielectric-filled waveguide. (a) Dispersion diagram, showing the asymmetric shift of the hyperbolic trajectory. Modes that were traveling at speed c remain so after the boost, indicated by the closely spaced open markers on the lightlike diagonal. (b) Lorentz phase for a wave traveling at speed c. No matter the rapidity of the boost, the x' axis still does and always will cut through the phase history of the wave rather than its future peaks.

backward-traveling waves accelerate, both approaching c in the limit of a boost with infinite rapidity. Overall, the new dispersion curve is asymmetric.

Recall that no boost can render a fast wave into a slow wave and vice versa. This is illustrated by the two closely spaced markers in Figure 7.7(a), both on the lightlike diagonal. These indicate a forward-traveling wave moving at speed c, a wave which remains on the lightlike contour after the boost. This is also shown on the Minkowski diagram of Figure 7.7(b); the wavefront passing through the origin remains halfway between the spatial and temporal axes both before and after the boost.

7.3.3 Phase Velocity and Group Velocity

Another way to look at these transformations is to plot the phase velocities as a function of frequency in a given observational reference frame, as illustrated in Figure 7.8(a). The phase velocity is always given by

$$v_p = \frac{\omega}{\beta} \tag{7.65}$$

and may be positive or negative (along with β depending on the direction of propagation). This ratio, if applied to the transformed frequency and propagation constant will be consistent with the velocity addition formula derived much earlier

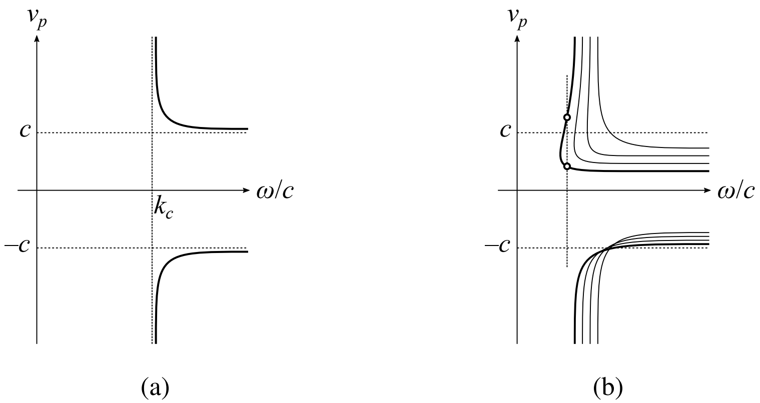


Figure 7.8 Phase velocity in a closed metallic waveguide. Positive and negative velocity correspond to forward and backward traveling waves, respectively. (a) Unfilled waveguide. (b) Dielectric-filled waveguide at different observational velocities. The open markers and dashed vertical line indicate a frequency for which both modes are forward-traveling.

in Chapter 3,

$$v_p' = \frac{\omega'}{\beta'} = \frac{v_p \cosh\zeta - c\sinh\zeta}{c\cosh\zeta - v_p\sinh\zeta} = \frac{v_p - c\tanh\zeta}{c - v_p\tanh\zeta} \tag{7.66}$$

For waves in an unfilled waveguide, the velocity spectrum shown never changes, though individual waves may move along these curves, even to the point of crossing the asymptote from positive velocity to negative.

The velocity for dielectric-filled waveguides, shown in Figure 7.8(b), is dependent on the motion of the internal medium. As was the case for plane waves in an unguided medium, some frequencies, given a sufficiently large boost, admit only waves traveling in one direction, such as those indicated by the two open markers and vertical dashed line on the plot. Dragging of these subluminal waves in a very rapidly moving medium has given them both a positive phase velocity.

What are we to make of the waves for which the phase velocity is greater than c? As we described before in Chapter 3, this is no violation of causality because the sinusoidal waves in this case are infinitely long in both space and time, and the phase velocity is merely an ensemble effect brought about by the interference of several plane waves acting together. It was promised that in all cases the propagation of energy and information would ultimately obey the relativistic speed limit. In waveguides, we now have a physical theory with which to test that assertion.

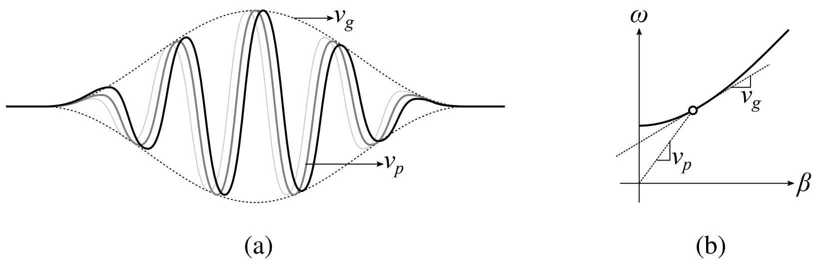


Figure 7.9 (a) Wave packet illustrating the difference between phase velocity and group velocity. (b) The phase velocity is given by the position of a point in the dispersion diagram, whereas group velocity is given by its slope.

Whereas the phase velocity ($v_p = \omega/\beta$) tracks the wavefronts in an infinitely long sinusoidal wave, the group velocity, given by

$$v_g = \frac{d\omega}{d\beta} \tag{7.67}$$

tracks the envelope of a wave packet, such as that in Figure 7.9(a), which can be more directly associated with the propagation of energy or information (as in a modulated waveform). The individual wavefronts within a packet may thread through it faster than c (and even in the opposite direction), but the speed at which the envelope of the wave packet translates must always be less than c.[7]

A Minkowski diagram showing the propagation of such a wave packet is given in Figure 7.10(a). One of the wavefronts has been marked using an open circle that skips forward rapidly as time progresses, indicated by the horizontal axes at times $t_1 \ldots t_2$. The extent of the wave packet is highlighted by the thick black line, which slides forward much more slowly.

It is possible to boost into a reference frame, even in an unfilled waveguide, in which the group velocity has changed signs, as illustrated in Figure 7.10(b). The worldline, ct', now leans rightward more than the envelope of the wave packet,

7 There are some unusual cases, typically associated with extreme levels of attenuation, in which the peak of an impulse may appear to move faster than c. However, it has been shown that these cases do not strictly correspond to the flow of energy or information — the peak of the much weaker and distorted output pulse is actually associated with the leading edge of the much wider input pulse. A more rigorous analysis surrounding the flow of energy and information shows that these are governed by yet another measure called the front velocity, and that this is always bounded by c [8].

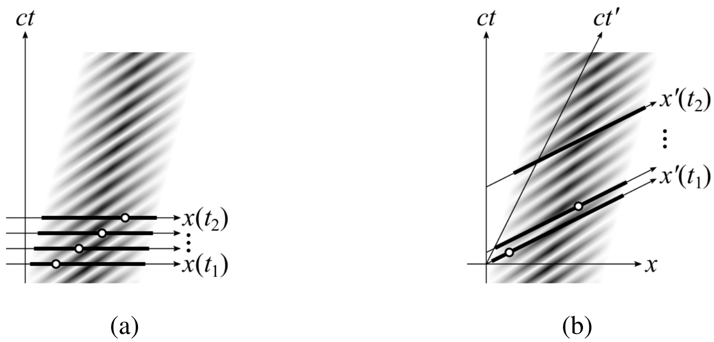


Figure 7.10 Minkowski diagram for a wave packet having $v_p > c$ and $v_g < c$. (a) One of the wave peaks is tracked through several instants in time with an open marker, while the extent of the packet is shown with the thick black line. (b) In a boosted reference frame, the phase velocity is still positive (and extremely fast), while the group velocity has changed sign, such that the packet is now sliding backwards (relative to the worldline, ct').

so that the packet itself appears to lag behind (i.e., the thick black lines slide left relative to this worldline). The phase velocity in this case is still directed forward, and extremely fast, as indicated by the open markers tagging a particular wavefront.

7.4 COAXIAL LINE

The closed metallic waveguides we have studied so far have cutoff frequencies which scale according to the cross-section dimensions, typically on the order of a wavelength (ridged waveguides are the only common forms of single-conductor geometries which can be said to have electrically small cross-section [1]). This makes them rather bulky as well as limited in bandwidth. There are other wave-guiding structures having more than one conductor that do work down to DC. These can be made electrically small compared to a wavelength, and in that regime are referred to as *transmission lines*. One of the most common forms is coaxial transmission line, shown in Figure 7.11. Although straight, air-filled coax is used sometimes over short distances, longer flexible cables must have some kind of dielectric insulator to support the center conductor mechanically.

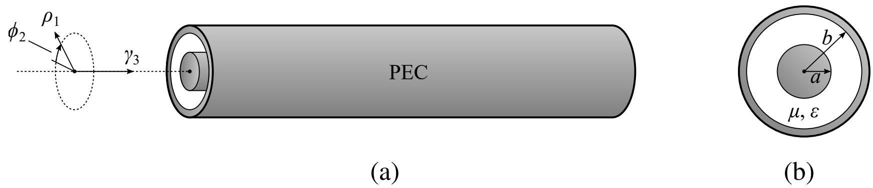


Figure 7.11 (a) Geometry of a coaxial transmission line and the cylindrical coordinate system that will be used to analyze it. (b) Detail of cross-section dimensions.

7.4.1 Transverse Mode

If the cross-section is electrically small, then longitudinal fields cannot be supported. A longitudinal polar field component would have to be zero at both the inner and outer conductors, and thus could only be nonzero in between if there was something on the order of half a wavelength between the inner and outer radii. A longitudinal axial field could potentially satisfy the differential boundary condition at both conductors, but since axial fields do not terminate on conductors, those representing a propagating wave must eventually form closed loops, requiring something on the order of a wavelength in circumference inside the dielectric. Neither of these conditions is consistent with the assumption that the lateral geometry is electrically small.

For these reasons, the formulas we derived for transverse axial and transverse polar modes do not apply for the dominant mode in the electrically small regime (they would apply for higher order modes when the structure is electrically large). Instead, our dominant mode solution will be entirely transverse. Such a mode is conventionally referred to as *transverse electric and magnetic* (TEM), but in present context may be called *transverse axial and polar* (TAP).

We may therefore conclude that the bivector field has only transverse components, no azimuthal dependency, and longitudinal dependency that is captured entirely by a sinusoidal phase term,

$$\mathcal{F} = \mathbf{f}_t(\rho) \cos\left(\varphi(t, z)\right) \tag{7.68}$$

Furthermore, in order to satisfy the lateral boundary conditions, the polar aspect of the bivector field must have no azimuthal component, while the axial aspect has no radial component. We may thus write

$$\mathbf{f}_t(\rho) = f_1 \rho_{10} + i f_2 \phi_{20} \tag{7.69}$$

Returning to (7.68), if we apply Maxwell's equation for a source-free region, we have

$$\blacksquare \mathcal{F} = (\blacksquare_{\mathbf{t}} \mathbf{f}_t) \cos\varphi - (\blacksquare_{\mathbf{z}} \varphi)\, \mathbf{f}_t \sin\varphi = 0 \tag{7.70}$$

This can only be satisfied for all phases φ if both terms are identically zero,

$$\blacksquare_{\mathbf{t}} \mathbf{f}_t = (\blacksquare_{\mathbf{z}} \varphi)\, \mathbf{f}_t = 0 \tag{7.71}$$

(Had we retained the $\sin\varphi$ and $\cos\varphi$ terms and not required the phase gradient to be zero, we would have found instead that $\blacksquare_{\mathbf{z}} \varphi \propto \cot\varphi$, an unbounded derivative which is nonphysical.) Let us start with the first term,

$$\blacksquare_{\mathbf{t}} \mathbf{f}_t = \blacksquare_{\mathbf{t}} \cdot \mathbf{f}_t + \blacksquare_{\mathbf{t}} \wedge \mathbf{f}_t = 0 \tag{7.72}$$

Therefore, after separating by grade,

$$\blacksquare_{\mathbf{t}} \cdot \mathbf{f}_t = 0 \tag{7.73a}$$

$$\blacksquare_{\mathbf{t}} \wedge \mathbf{f}_t = 0 \tag{7.73b}$$

We can use the spacetime splits given in (D.64), and apply to them the cylindrical forms of the divergence and curl operations from Appendix B. Since we have limited both the operator and the field quantity here to certain transverse terms only, the resulting expressions are relatively simple. The divergence equation becomes

$$(\blacksquare_{\mathbf{t}} \cdot \mathbf{f}_t)\, \gamma_0 = (\blacksquare_{\mathbf{t}} \cdot (f_1 \rho_{10} + i f_2 \phi_{20}))\, \gamma_0 = \boldsymbol{\nabla}_t \cdot (f_1 \rho_{10}) + \boldsymbol{\nabla}_t \times (f_2 \phi_{20}) \tag{7.74a}$$

$$= \frac{1}{\rho} \frac{\partial}{\partial \rho} (\rho f_1) + \frac{1}{\rho} \frac{\partial}{\partial \rho} (\rho f_2)\, \sigma_{30} = 0 \tag{7.74b}$$

From this we may conclude

$$f_1 = \frac{F_1}{\rho} \tag{7.75a}$$

$$f_2 = \frac{F_2}{\rho} \tag{7.75b}$$

for scalar constants F_1 and F_2. The curl equation is automatically satisfied with the constraints we have predetermined,

$$(\blacksquare_{\mathbf{t}} \wedge \mathbf{f}_t)\, \gamma_0 i = (\blacksquare_{\mathbf{t}} \wedge (f_1 \rho_{10} + i f_2 \phi_{20}))\, \gamma_0 i \tag{7.76a}$$

$$= \boldsymbol{\nabla}_t \cdot (f_2 \phi_{20}) - \boldsymbol{\nabla}_t \times (f_1 \rho_{10}) = 0 \tag{7.76b}$$

It remains to determine the phase gradient from

$$(\blacksquare_{\mathbf{z}}\varphi)\,\mathbf{f}_t = L\mathbf{f}_t = 0 \tag{7.77}$$

As usual, L must be a constant lying in the γ_{30} plane. We may use the usual parameterization,

$$\blacksquare_{\mathbf{z}}\varphi = L \tag{7.78a}$$

$$\varphi = K \cdot x \tag{7.78b}$$

where

$$L = k\,(\gamma_0 + \gamma_3) \tag{7.79a}$$

$$K = k\left(n^{-1}\gamma_0 + \gamma_3\right) \tag{7.79b}$$

We have used $\beta = k = n\omega/c$ for the propagation constant here since L is required to be a null vector ($L^2 = 0$), lest we could invert it in (7.77) and find that $\mathbf{f}_t = 0$. Instead, substituting the parameters we have determined into (7.77), we have

$$L\mathbf{f}_t = k\,(\gamma_0 + \gamma_3)\,\tfrac{1}{\rho}\,(F_1\rho_{10} + iF_2\phi_{20}) = 0 \tag{7.80a}$$

$$\therefore (\gamma_0 + \gamma_3)\,(F_1\rho_1\gamma_0 + F_2\rho_1\gamma_3) = 0 \tag{7.80b}$$

$$(F_1 - F_2)\,(1 + \sigma_3) = 0 \tag{7.80c}$$

from which we may immediately conclude that $F_1 = F_2$.

The complete transverse mode solution for an electrically thin (small in cross-section) coaxial line is therefore given by

$$\mathcal{F} = \tfrac{F_1}{\rho}\,(\rho_{10} + i\phi_{20})\cos(K \cdot x) \tag{7.81}$$

and is plotted in Figure 7.12.

7.4.2 Parasitic Transverse Polar Mode

Coaxial line is almost never (intentionally) operated above its single-mode frequency range, but it is useful to determine at least the cutoff frequency of the next higher-order mode so that the structure's useful bandwidth can be quantified. The first nondominant mode is a transverse polar mode, whose derivation is the same as that for cylindrical waveguide up to the application of boundary conditions,

$$\mathbf{f}_z = if_z\sigma_3 \tag{7.82a}$$

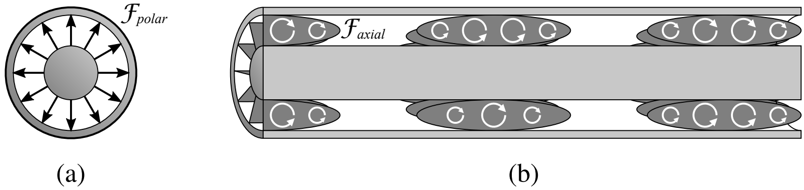


Figure 7.12 Faraday field components for the transverse mode of an electrically thin coaxial line. (a) End view showing polar bivector components. (b) Side cutaway view showing axial bivector components.

$$f_z = k_c^2 \left(F_1 J_n \left(k_c \rho\right) + F_2 Y_n \left(k_c \rho\right)\right) \cos\left(n(\phi + \psi)\right) \tag{7.82b}$$

We previously excluded Y_n, the Bessel function of the second kind, since it became infinite at the origin ($\rho = 0$). In this case, the origin is outside the domain of the bivector field (being covered by the center conductor) so the function of the second kind is admissible.

The boundary conditions constrain the radial slope of the field to be zero at $\rho = a$ and $\rho = b$. Therefore,

$$F_1 J_n' \left(k_c a\right) + F_2 Y_n' \left(k_c a\right) = 0 \tag{7.83a}$$

$$F_1 J_n' \left(k_c b\right) + F_2 Y_n' \left(k_c b\right) = 0 \tag{7.83b}$$

This system of equations can be written in matrix form as

$$\begin{pmatrix} J_n' \left(k_c a\right) & Y_n' \left(k_c a\right) \\ J_n' \left(k_c b\right) & Y_n' \left(k_c b\right) \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \end{pmatrix} = 0 \tag{7.84}$$

and can be satisfied for nontrivial F_1 and F_2 only when

$$\det \begin{pmatrix} J_n' \left(k_c a\right) & Y_n' \left(k_c a\right) \\ J_n' \left(k_c b\right) & Y_n' \left(k_c b\right) \end{pmatrix} = J_n' \left(k_c a\right) Y_n' \left(k_c b\right) - J_n' \left(k_c b\right) Y_n' \left(k_c a\right) = 0 \tag{7.85}$$

The solutions to this transcendental equation determine the cutoff wavenumbers, k_c, for various high-order transverse polar modes. In general, this requires numerical computation, but the first higher-order mode with which we are most interested can be very closely approximated as

$$k_c \approx \frac{2}{a + b} \tag{7.86}$$

Given that $k_c = 2\pi/\lambda_g$, where λ_g is the wavelength at cutoff inside the dielectric medium, this result can be recast as follows,

$$\lambda_g = \frac{2\pi}{k_c} = \pi(a+b) \tag{7.87}$$

In other words, coaxial line is single-moded so long as the average circumference inside the dielectric is smaller than a wavelength [1] — a condition alluded to in the assumptions made at the beginning of Section 7.4.1.

References

[1] M. A. Morgan, *Principles of RF and Microwave Design*. Norwood, MA: Artech House, November 2019.

[2] S. Ramo, J. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics*. New York: Wiley, 1984.

[3] D. M. Pozar, *Microwave Engineering*, 4th ed. New York: Wiley, 2011.

[4] R. Haberman, *Elementary Applied Partial Differential Equations: with Fourier Series and Boundary Value Problems*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1987.

[5] G. Arfken and H. Weber, *Mathematical Methods for Physicists*, 4th ed. San Diego, CA: Academic Press, 1995.

[6] W. E. Boyce and R. C. DiPrima, *Elementary Differential Equations and Boundary Value Problems*, 6th ed. New York: Wiley, 1997.

[7] Wikipedia. (2024, January) Bessel function. https://en.wikipedia.org/wiki/Bessel_function.

[8] P. W. Milonni, *Fast Light, Slow Light and Left-Handed Light*. New York: Taylor & Francis, 2005.

Chapter 8

Network Analysis

The first half of this book focused on certain fundamental physical laws in the broadest context — electromagnetics and relativity in either the vacuum of space or a homogeneous material environment, with at most one planar interface. Starting in Chapter 7, we began to refine those laws in the presence of engineered boundary conditions (e.g., waveguides) whose purpose is to control the Faraday field for practical applications. We continue that refinement in this chapter with the analysis of structures having electrically small interfaces. This will ultimately lead us to the concepts of lumped impedance and network parameters, formalisms within which most day-to-day electrical engineering takes place.

8.1 INTEGRAL FORMS

Thus far, our exposition of electromagnetic laws in four-dimensional spacetime has been limited to purely algebraic and differential forms; the process of integration has not yet been discussed. It is useful to introduce this operation now as it will be pivotal to the development of the scalar quantities (i.e., voltage and current) that allow us to describe circuit characteristics more simply than we can using fully fleshed bivector fields.

8.1.1 Directed Integration

The most general form of an integral in spacetime is

$$I = \int_{V_k} \mathcal{M} d^k x \mathcal{N} \tag{8.1}$$

for general multivector fields $\mathcal{M}$ and $\mathcal{N}$ integrated over the k-volume V_k. The differential element, $d^k x$, represents an oriented k-blade[1] that can be expanded via the wedge product,

$$d^k x = dx_1 \wedge dx_2 \wedge \cdots \wedge dx_k \tag{8.2}$$

where $dx_\mu = |dx|\gamma'_\mu$ for some basis vectors γ'_μ that are locally defined and orthonormal (in the Minkowski sense). We may also write

$$d^k x = |dx|^k O_k \tag{8.3}$$

where the scalar $|dx| \to 0$ and the directional orientation is encoded in the parameter O_k having unit norm. The juxtaposition of elements within the integral is interpreted in the usual way as a geometric product, and both $\mathcal{M}$ and $\mathcal{N}$ are included since the commutivity rules of spacetime algebra suggest that both left and right integrands should be allowed.

It is important not to lose sight of the directionality of integration. As an example, the directed integration of spacetime around any k-volume having no boundary (e.g., a closed loop or surface) does not measure the extent of that region (as in the circumference or area) but instead is identically zero,

$$\oint_{V_k} d^k x = 0 \tag{8.4}$$

This is illustrated in Figure 8.1(a) for a closed loop. The directed integral around that loop is effectively an infinite summation of vectors that follow the contour of the loop around a complete circuit, as shown in Figure 8.1(b), and is always zero. To recover the circumference, C, we must perform a scalar integration by multiplying the inverse of the orientation parameter,

$$C = \oint_{L_1} \left|d^1 x\right| = \oint_{L_1} d^1 x O_1^{-1} \tag{8.5}$$

as shown in Figure 8.1(c).

Certain special forms apply to the differentials of particular grades. For example, the grade-4 differential is always given by

$$d^4 x = \pm i |dx|^4 \tag{8.6}$$

1 Appreciate the difference between a k-volume, which is a k-dimensional entity defining a region with clear boundaries and shape, but not necessarily an orientation, and a k-blade, which in contrast is a k-dimensional entity having clear orientation and magnitude, but no clear shape or boundaries, as previously described.

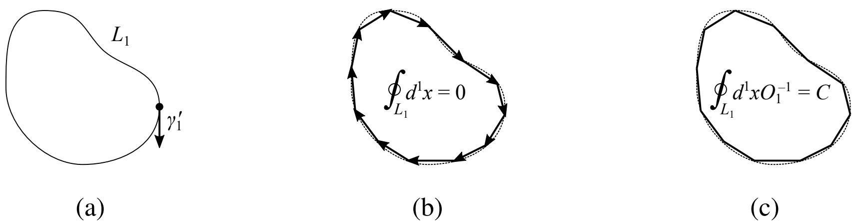


Figure 8.1 Effect of directionality on integration. (a) A closed loop of dimension $k = 1$. (b) A directed integral around the loop is zero. (c) A scalar integral around the loop is equal to the circumference.

since the wedge product of any four independent basis vectors is always the unit pseudoscalar, i, to within a sign determined by sequence. Similarly, the grade-3 differential may be written

$$d^3x = iN|dx|^3 \tag{8.7}$$

where N is a unit normal vector to the corresponding three-volume. Note the similarity of this equation to (6.94); the direction of the unit normal depends on the orientation of d^3x.

8.1.2 The Fundamental Theorem

The fundamental theorem of calculus in spacetime algebra, also called the *boundary theorem*, states that the directed integral of a derivative throughout a k-volume is equal to the directed integral of the (undifferentiated) field on the closed boundary of that k-volume,

$$\int_{V_k} \mathcal{M} d^k x \Box_{V_k} \mathcal{N} = \oint_{\partial V_k} \mathcal{M} d^{k-1} x \mathcal{N} \tag{8.8}$$

(noting that the tangential derivative $\Box_{V_k}$ may apply to either the right or the left, that is to $\mathcal{M}$ or $\mathcal{N}$, or both). This is an extraordinarily general theorem, embodying the classical Kelvin-Stokes theorem, the divergence theorem, Cauchy's theorems, Green's theorems, and the more familiar fundamental theorem of (scalar) calculus, as well as numerous other results without names [1, 2].

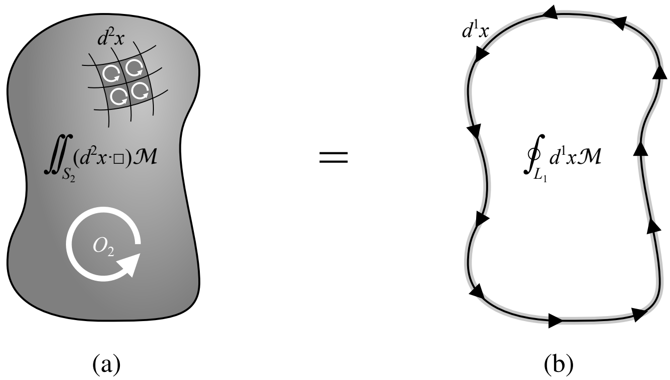


Figure 8.2 The fundamental theorem applied to an open two-dimensional surface. (a) The interior 2-volume (surface) has counterclockwise-oriented elements of area, d^2x. (b) The 1-volume boundary (loop) inherits that directionality, with counterclockwise-directed length elements, d^1x.

It should be noted that the combination of the oriented differential with the tangential gradient operator may be reduced to a dot product form as follows,[2]

$$d^kx\square_{V_K} = d^kx\square_{d^kx} = d^kx\left(d^kx\right)^{-1}\left(d^kx\cdot\square\right) = d^kx\cdot\square \tag{8.9}$$

Additionally omitting the left integrand allows us to write the fundamental theorem in a somewhat simpler form that is still quite useful in most cases,

$$\int_{V_k}\left(d^kx\cdot\square\right)\mathcal{M} = \oint_{\partial V_k} d^{k-1}x\mathcal{M} \tag{8.10}$$

The boundary integration inherits its directionality from the orientation of the interior integration as shown in Figure 8.2.

8.1.3 Maxwell's Equation in Integral Form

Consider the integral of the Faraday field over a three-dimensional region of spacetime. According to (8.10),

$$\iiint_V\left(d^3x\cdot\square\right)F = \oiint_S d^2xF \tag{8.11}$$

2 This equivalence depends on the fact that d^kx is always written as a single wedge product, and thus is a simple bivector when $k=2$. See Section D.4 for more details.

The integrand on each side represents a product of two bivectors, having scalar, bivector, and pseudoscalar parts. We may write these as separate equations,

$$\iiint_V \left(d^3x \cdot \Box\right) \cdot F = \oiint_S d^2x \cdot F \tag{8.12a}$$

$$\iiint_V \left[d^3x \cdot \Box, F\right] = \oiint_S \left[d^2x, F\right] \tag{8.12b}$$

$$\iiint_V \left(d^3x \cdot \Box\right) \wedge F = \oiint_S d^2x \wedge F \tag{8.12c}$$

Recall that square brackets denote the commutator product. The first, scalar equation can be simplified using the identity (D.139) along with the homogeneous Maxwell's equation ($\Box \wedge F = 0$),

$$\iiint_V \left(d^3x \cdot \Box\right) \cdot F = \iiint_V d^3x \cdot (\Box \wedge F) = 0 \tag{8.13a}$$

$$\therefore \oiint_S d^2x \cdot F = 0 \tag{8.13b}$$

This integral equation embodies the homogeneous Maxwell's equations: Faraday's law and the absence of magnetic charge.

To give an example, suppose that the surface S is an entirely spatial enclosure (in a given reference frame); it is thus everywhere orthogonal to γ_0 and must have its basis in $i\sigma_k$. We then have

$$\oiint_S d^2x \cdot F = \oiint_S d^2x \cdot \left(\tfrac{1}{c}\mathbf{E} + i\mathbf{B}\right) = \oiint_S \mathbf{B} \cdot d\mathbf{S} = 0 \tag{8.14}$$

which is Maxwell's equation for the absence of magnetic charge. In contrast, if S comprises two connected parts, one which is spatial, and the other which has a temporal component (that is, its basis is in σ_k), then we may write

$$\oiint_S d^2x \cdot F = \frac{1}{c}\iint_{S_1} d^2x \cdot \mathbf{E} + \iint_{S_2} d^2x \cdot (i\mathbf{B}) = 0 \tag{8.15}$$

Note that we have replaced the closed-surface integral on S with two open-surface integrals on the conjoined regions S_1 and S_2. Now let us differentiate with respect to time,

$$\frac{1}{c}\frac{\partial}{\partial t}\iint_{S_1} d^2x \cdot \mathbf{E} + \frac{\partial}{\partial t}\iint_{S_2} d^2x \cdot (i\mathbf{B}) = 0 \tag{8.16}$$

By assumption, the region S_1 has its basis in $\sigma_k = \gamma_k\gamma_0$, and thus the double-integral over $\mathbf{E}$ comprises one integration with respect to time. Differentiation simply cancels that integral in accordance with the classic fundamental theorem of calculus, leaving behind a linear integral over a (closed) path that is its boundary with S_2,

$$\oint_{L_1} \mathbf{E} \cdot d\mathbf{l} + \frac{\partial}{\partial t}\iint_{S_2} \mathbf{B} \cdot d\mathbf{S} = 0 \tag{8.17}$$

which is Faraday's law.

The pseudoscalar equation (8.12c) may also be simplified with the help of (8.7) and the duality of dot and wedge products,

$$\iiint_V \left(d^3x \cdot \Box\right) \wedge F = \iiint_V |dx|^3 \left((iN) \cdot \Box\right) \wedge F \tag{8.18a}$$

$$= \iiint_V |dx|^3 N \wedge \left((\Box \cdot F)\, i\right) = \iiint_V |dx|^3 N \wedge (Ji) \tag{8.18b}$$

$$= \iiint_V |dx|^3 (iN) \wedge J = \iiint_V d^3x \wedge J \tag{8.18c}$$

$$\therefore \oiint_S d^2x \wedge F = \iiint_V d^3x \wedge J \tag{8.18d}$$

This is the inhomogeneous Maxwell's equation, or Gauss-Ampère law.

The two integral laws we have thus derived, (8.13b) and (8.18d), may be written as one using grade projection [3],

$$\oiint_S \left\langle d^2xF \right\rangle_i = \iiint_V \left\langle d^3xJ \right\rangle_i \tag{8.19}$$

where $\langle \mathcal{M} \rangle_i$ represents the Lorentz-invariant part (the scalar and pseudoscalar elements) of $\mathcal{M}$.

These laws may be adapted to the macroscopic Faraday field if the vector elements of the oriented differential d^kx are scaled by the proper index of refraction according to the projection and rejection with respect to the four-velocity of the medium,

$$dx'_\mu = n^{-1}\left(dx_\mu\right)_U + \left(dx_\mu\right)_{\not{U}} \tag{8.20}$$

Table 8.1
Macroscopic Maxwell's Equations in Differential and Integral Form

Name	Differential Form	Integral Form
Homogeneous equation	$\blacksquare \wedge \mathcal{F} = 0$	$\oiint_S \mathbf{d}^2x \cdot \mathcal{F} = 0$
Inhomogeneous equation	$\blacksquare \cdot \mathcal{F} = \mathcal{J}$	$\oiint_S \mathbf{d}^2x \wedge \mathcal{F} = \iiint_V \mathbf{d}^3x \wedge \mathcal{J}$
Combined equation	$\blacksquare\mathcal{F} = \mathcal{J}$	$\oiint_S \left\langle \mathbf{d}^2x\mathcal{F}\right\rangle_i = \iiint_V \left\langle \mathbf{d}^3x\mathcal{J}\right\rangle_i$

In the rest frame, this merely replaces the light-speed scaling factor of the temporal component with the phase velocity in the medium (i.e., $ct \to v_pt$). Such differentials will be written with a boldface $\mathbf{d}$ as below,

$$\mathbf{d}^kx = dx'_1 \wedge dx'_2 \wedge \cdots \wedge dx'_k \tag{8.21}$$

The macroscopic Maxwell's equation then becomes

$$\oiint_S \left\langle \mathbf{d}^2x\mathcal{F}\right\rangle_i = \iiint_V \left\langle \mathbf{d}^3x\mathcal{J}\right\rangle_i \tag{8.22}$$

The complete macroscopic field equations are thus shown in Table 8.1 in both differential and integral form.

8.2 COMPACT PORTS

We made an assumption when deriving the dominant mode of coaxial line in Section 7.4 that the cross-section was small compared to the wavelength of operation. This condition could not be satisfied by the rectangular and circular waveguide geometries whose lateral dimensions were inexorably linked to the size of the waves they carried. Coaxial line was thus classified as a transmission line, a waveguiding structure that is potentially very long but skinny compared to the waves it supports [4–6].

8.2.1 Terminal Parameters

One of the consequences of an electrically small port is that the Faraday field outside of the conductors is conservative; any closed contour integral confined to such

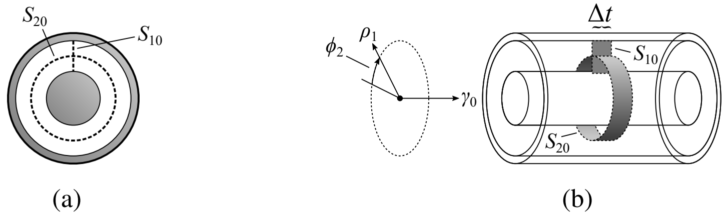


Figure 8.3 Integration volumes (2-blades) for determination of terminal parameters. (a) Cross-section view. (b) Extent on temporal axis (not the spatial longitudinal axis). For present purposes, we use a delta function in time and allow $\Delta t \to 0$.

a compact region shall vanish unless that contour encloses an electrically dense source charge or current (in classical electromagnetics, it is said that a line integral of this type depends only on its endpoints, not the path taken). We may thus expect to distill all there is to know about a wave propagating on a transmission line down to a couple of scalar quantities by aggregating those sources.

To that end, let us integrate the polar field from one port terminal to the other, and the axial field (which does not terminate) around a closed loop encircling one of the conductors in the cross-section plane. The conservativeness of the field in an electrically small region guarantees that the specific path of the integral does not matter, so we choose the radial and concentric paths shown in Figure 8.3 for simplicity. As bivectors, the simplest (i.e., scalar) result is obtained by 2-blade integrations, so we include a temporal component as well. But since our interest is in the terminal parameters at a particular instant and at a particular longitudinal position on the line, the extent of the temporal region may be arbitrarily small, and weighted by a delta function. The integrals are thus evaluated as follows,

$$V = \iint_{S_{10}} \mathcal{F} \cdot \mathbf{d}^2 x = \int_a^b (\mathcal{F} \cdot \rho_{10})\, d\rho = (F_1 \cos\varphi) \int_a^b \rho^{-1} d\rho \tag{8.23a}$$

$$= F_1 \ln\left(\tfrac{b}{a}\right) \cos\varphi \tag{8.23b}$$

$$i\eta I = \iint_{S_{20}} \mathcal{F} \wedge \mathbf{d}^2 x = \int_0^{2\pi} \mathcal{F} \wedge \phi_{20} \rho d\phi = i\,(F_1 \cos\varphi) \int_0^{2\pi} d\phi \tag{8.23c}$$

$$= i2\pi F_1 \cos\varphi \tag{8.23d}$$

Note that we have used the dot product to isolate the polar field, and the wedge product to isolate the axial field. Drawing from our experience in classical electromagnetics, we have associated these results with the voltage (V) and current (I), respectively. The characteristic impedance of the transmission line is given by their ratio, and should be recognizable to anyone familiar with coax,

$$Z_0 = \frac{\eta}{2\pi} \ln\left(\frac{b}{a}\right) \tag{8.24}$$

8.2.2 Phasor Amplitudes

Thus, while we have built our theory in this book on the atypical foundation of special relativity, we have nevertheless arrived at the familiar territory of propagating voltage and current waves. Rewriting the results of (8.23) in a condensed form, we have

$$V = v_0 \cos\varphi \tag{8.25a}$$

$$I = i_0 \cos\varphi \tag{8.25b}$$

or, more generally for the combined effects of forward and backward traveling waves,

$$V = v_0^+ \cos\varphi + v_0^- \cos\psi \tag{8.26a}$$

$$I = i_0^+ \cos\varphi + i_0^- \cos\psi \tag{8.26b}$$

where

$$v_0^+ = Z_0 i_0^+ \tag{8.27a}$$

$$v_0^- = Z_0 i_0^- \tag{8.27b}$$

$$\varphi = K \cdot x = \omega t - kz \tag{8.27c}$$

$$\psi = K^\dagger \cdot x = \omega t + kz \tag{8.27d}$$

The utility of voltage and current as scalar tracers of a guided Faraday wave lies in the fact that these quantities are preserved at junctions between transmission lines and at lumped (electrically compact) elements. For example, the charge density may change suddenly when passing from one transmission line to another, as shown in Figure 8.4, but the net voltage is conserved. This makes it easier to write equations that govern the interaction between waves on either side of the boundary, and thereby to derive reflection and transmission coefficients based solely

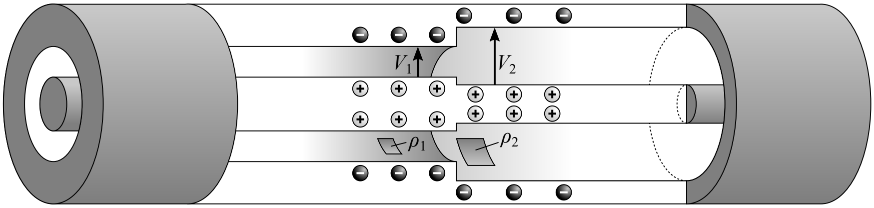


Figure 8.4 Conservation of voltage across an impedance step in coaxial transmission line. The charge density is discontinuous, $\rho_1 \neq \rho_2$, but the voltage is constant, $V_1 = V2$, sufficiently close to the boundary.

on the impedance properties of the transmission lines, rather than on their specific geometries. However, to do this in the usual way requires a mathematical trick exploiting the behavior of complex numbers known as the *phasor transform* [4].

In light of our newfound physical interpretation of the unit pseudoscalar i as an oriented four-volume in spacetime, we must be cautious about reintroducing it here in a totally artificial manner. Expressing voltage and current waves as a complex phasor amplitude should not imply that these scalar elements have somehow gained a four-dimensional aspect to them. Let us therefore write the total voltage and current, temporarily at least, in terms of the object j, for which $j^2 = -1$,

$$V = v_0^+ \mathrm{Re}_j \left\{ e^{j(\omega t - kz)} \right\} + v_0^- \mathrm{Re}_j \left\{ e^{j(\omega t + kz)} \right\} \tag{8.28a}$$

$$I = i_0^+ \mathrm{Re}_j \left\{ e^{j(\omega t - kz)} \right\} - i_0^- \mathrm{Re}_j \left\{ e^{j(\omega t + kz)} \right\} \tag{8.28b}$$

(The sign of the current amplitude for a backward traveling wave has been inverted due to the change in reference direction.) We have not yet fully defined what j is, only that it squares to minus one (and since all other terms in this equation are scalars, there is no question that they commute with j). The notation $\mathrm{Re}_j \{.\}$ defines the real part of an expression with respect to j, but not with respect to any other imaginary units like i.

Unlike i, which had a clear physical interpretation, j has been introduced here purely for mathematical convenience. The object j is therefore truly imaginary in a way that the pseudoscalar i is not. Any factor of j that remains in the final solution to a problem must be considered the leftovers of a mathematical trick, to be discarded as something which does not belong to the true result.

As the reader is no doubt aware, it is customary to drop the common exponential factor in $j\omega t$ and write the phasor equations,

$$V' = v^+ e^{-jkz} + v^- e^{+jkz} \tag{8.29a}$$

$$I' = i^+ e^{-jkz} - i^- e^{+jkz} \tag{8.29b}$$

where V', for example, is the complex phasor amplitude of V, and the true signal can always be recovered from it as follows,

$$V = \mathrm{Re}_j \left\{ V' e^{j\omega t} \right\} \tag{8.30}$$

8.2.3 Reflection Coefficient and Impedance

The continuity of voltage and current at junctions between transmission lines ensures that the ratio of the net voltage and current resulting from the sum for forward and backward-traveling waves at a step discontinuity, such as that shown in Figure 8.4, must be equal to the characteristic impedance of the loading transmission line, Z_L. In other words,

$$\frac{V'}{I'} = \frac{v^+ e^{-j\theta} + v^- e^{+j\theta}}{i^+ e^{-j\theta} - i^- e^{+j\theta}} = Z_0 \frac{i^+ e^{-j\theta} + i^- e^{+j\theta}}{i^+ e^{-j\theta} - i^- e^{+j\theta}} = Z_L \tag{8.31}$$

where the discontinuity occurs at $kz = \theta$. The same equation would result from a lumped termination or load having impedance Z_L. The reflection coefficient is found by solving for the ratio of backward and forward-traveling (phasor) wave amplitudes on the incident line,

$$\Gamma = \frac{v^-}{v^+} = \frac{i^-}{i^+} = \frac{z_L - 1}{z_L + 1} e^{-j2\theta} \tag{8.32}$$

where $z_L = Z_L/Z_0$ is the normalized impedance of the load relative to the characteristic impedance of the incident transmission line. Typically, we measure the reflection coefficient directly at the discontinuity ($\theta = 0$),

$$\Gamma = \frac{z_L - 1}{z_L + 1} \tag{8.33}$$

This can be inverted to express the impedance in terms of the reflection coefficient,

$$z_L = \frac{1+\Gamma}{1-\Gamma} \tag{8.34}$$

Or, if we prefer, we can write the effective impedance seen at a standoff distance $kz = \theta$ from the discontinuity in terms of the load impedance,

$$z_{in} = \frac{1+\Gamma e^{-j2\theta}}{1-\Gamma e^{-j2\theta}} = \frac{1+\frac{z_L-1}{z_L+1}e^{-j2\theta}}{1-\frac{z_L-1}{z_L+1}e^{-j2\theta}} \tag{8.35a}$$

$$= \frac{z_L+1+(z_L-1)\,e^{-j2\theta}}{z_L+1-(z_L-1)\,e^{-j2\theta}} \tag{8.35b}$$

$$= \frac{z_L\left(e^{j\theta}+e^{-j\theta}\right)+\left(e^{j\theta}-e^{-j\theta}\right)}{\left(e^{j\theta}+e^{-j\theta}\right)+z_L\left(e^{j\theta}-e^{-j\theta}\right)} \tag{8.35c}$$

$$= \frac{z_L\cos\theta + j\sin\theta}{\cos\theta + jz_L\sin\theta} = \frac{z_L + j\tan\theta}{1+jz_L\tan\theta} \tag{8.35d}$$

These calculations are no doubt second nature to students and professionals of microwave engineering [4–7] — as are the difficulties encountered when applying them sequentially, say, to develop the combined impedance of series and parallel elements connected at intervals along a transmission line. Equations (8.32) to (8.35) are all *Möbius transformations* [8], rational functions having the following general form

$$f(x) = \frac{ax+b}{cx+d} \tag{8.36}$$

The Smith chart is effectively a nomographic tool for performing Möbius transformations visually.

The general nonlinearity of Möbius transformations makes them cumbersome in complex calculations. I myself have carried out layered Möbius transformations nested to the seventh order for the sake of deriving a filter's transfer function [9], a tedious exercise I care not to repeat! Arsenovic has shown [10] that spacetime algebra provides a better way, which will be detailed in Section 8.3.[3]

3 Some minor differences between Arsenovic's results and our own will be observed due to our choice of metric signature.

8.3 A NEW LANGUAGE FOR NETWORK ANALYSIS

As it happens, we have seen Möbius transformations elsewhere in this book. The Lorentz transformations, whether boosts, proper rotations, or some combination, are Möbius transformations, albeit a simple kind where the coefficient of x in the denominator of (8.36) is $c = 0$. A somewhat more traditional Möbius transformation may be recognized in the velocity addition formula,

$$v' = \frac{v \cosh \zeta - c \sinh \zeta}{c \cosh \zeta - v \sinh \zeta} \tag{8.37}$$

Since spacetime algebra is specially suited for performing these kinds of calculations in special relativity, it should not be surprising that it also furnishes us with precisely the toolkit we need to perform these same kinds of calculations more simply in the context of network analysis. To leverage it, we must first transfer the impedance information into a higher-dimensional space, using an operation called *stereographic projection*. The rotors needed to perform common network analysis operations will then be described in Section 8.4.

8.3.1 Stereographic Projection

A key feature of all Möbius transformations in complex numbers is that they can be shown equivalent to the following sequence of geometric operations:

1. Stereographic projection from the complex plane onto a sphere;
2. Rotation and translation of that sphere;
3. Reverse projection from the sphere back onto the complex plane.

We shall soon write down a formula for the stereographic projection in step 1 (and its inverse in step 3). The intervening spherical rotation, representing as many Möbius transformations as desired, may then be implemented in the second step using the two-sided rotor transformation introduced in Section 5.3. While there is some obvious overhead associated with the projection operations, concatenation of multiple Möbius transformations as a product of rotors is a much simpler task than nesting the nonlinear formula (8.36) directly. The intermediate spherical space that we use is generally referred to as the *Riemann sphere*.

In order to implement stereographic projection in the language of spacetime algebra, we must define a vector basis. I will opt not to use γ_μ for this, as it might lead to confusion with the physical spacetime dimensions. Let us instead call these

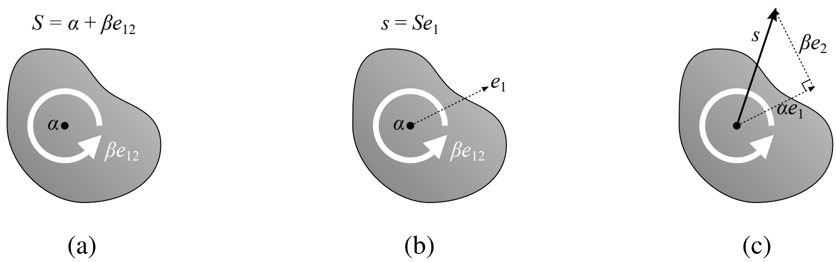


Figure 8.5 Definition of a complex number in terms of spacetime vectors. (a) Paravector $S = \alpha + \beta e_{12}$. (b) Reference direction e_1. (c) Vector $s = Se_1 = \alpha e_1 + \beta e_2$. Note that e_1, e_2, and s all lie in the e_{12} plane.

purely abstract dimensions e_μ, having the same metric signature as before,

$$e_\mu^2 = \begin{cases} +1 & \mu = 0 \\ -1 & \mu = 1, 2, 3 \end{cases} \tag{8.38}$$

A complex number can be written as the sum of two parts: one a scalar, and the other a coefficient multiplied by a suitably chosen imaginary unit,

$$S = \alpha + j\beta \tag{8.39}$$

We can select almost anything we wish from the vector basis and its higher-grade products as the imaginary unit, so long as it squares to minus one. Let us choose the bivector $j = e_{12}$; therefore,

$$S = \alpha + \beta e_{12} \tag{8.40}$$

By choosing a bivector imaginary unit, we have written this in a form that resembles the paravectors we derived earlier via the spacetime split. Such a paravector is illustrated graphically in Figure 8.5(a). We may convert this to a proper vector in the same way that a paravector was reformed into a spacetime four-vector, namely by multiplying it with a chosen reference axis, e_1, as in Figure 8.5(b, c),

$$s = Se_1 = \alpha e_1 + \beta e_2 \tag{8.41}$$

Now we are ready to map this vector onto the Riemann sphere using stereographic projection. The process is illustrated in Figure 8.6. The Riemann sphere is

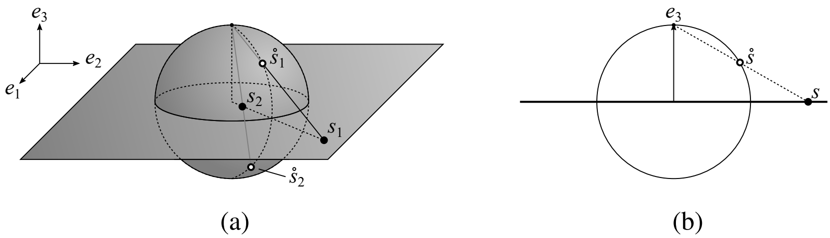


Figure 8.6 (a) Stereographic projection from the e_{12} plane onto the Riemann sphere in e_{123}. Points on the plane, s_k, are marked with a solid black circle, whereas their projected points on the sphere are labeled $\mathring{s}_k$ and marked with an open circle. (b) Cross-section cut.

situated at the origin where it is cut through at the equator by the complex plane. A straight line is drawn from the northern pole of the sphere to a point, s, on the complex plane. The intersection of that line with the surface of the sphere is the projected point, $\mathring{s}$. Mapping is one-to-one, with all points from inside the unit circle arriving in the lower hemisphere, and those outside the unit circle lying in the upper hemisphere. The northern pole represents the complex point at infinity (in this way, we map even infinity to a local point with finite coordinates, thus avoiding the singularities that sometimes crop up when using immittance parameters).

We may write the point $\mathring{s}$ as follows,

$$\mathring{s} = e_3 + \lambda \left(s - e_3\right) \tag{8.42}$$

for some scalar λ, and since $\mathring{s}$ is on the unit sphere (with a negative spatial signature), we have

$$\mathring{s}^2 = e_3^2 + \lambda^2 \left(s^2 + e_3^2\right) - 2\lambda e_3^2 = -1 \tag{8.43a}$$

$$-1 + \lambda^2 \left(s^2 - 1\right) + 2\lambda = -1 \tag{8.43b}$$

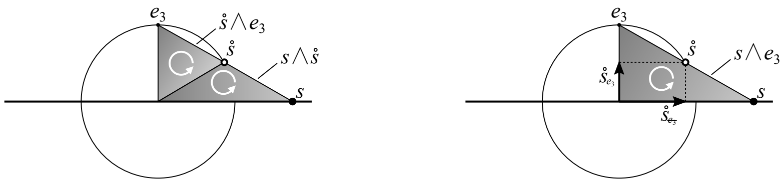
$$\lambda = \frac{2}{1 - s^2} = \frac{2}{1 + |s^2|} \tag{8.43c}$$

Thus,

$$\mathring{s} = e_3 + 2\frac{s - e_3}{1 - s^2} = \left(\frac{s^2 + 1}{s^2 - 1}\right) e_3 - \left(\frac{2}{s^2 - 1}\right) s \tag{8.44}$$

We shall also need to reverse the process, thereby expressing s in terms of $\mathring{s}$. Equation (8.44) is not easily inverted, but we may recognize that

$$s \wedge \mathring{s} + \mathring{s} \wedge e_3 = s \wedge e_3 \tag{8.45}$$



(a) (b)

Figure 8.7 Equality of (a) $s \wedge \mathring{s} + \mathring{s} \wedge e_3$ and (b) $s \wedge e_3$, used for reverse projection from the Riemann sphere back down onto the complex plane. (Technically, the triangles drawn here represent only half the area encoded in the relevant bivectors, but the equality is still valid because this common factor cancels out.) Note also the colinearity of e_3 with $\mathring{s}_{e_3}$, and of $\mathring{s}_{\not e_3}$ with s.

as shown in Figure 8.7. Then, if we expand $\mathring{s}$ in terms of its projection and rejection from e_3, we can solve for s,

$$s \wedge \left(\mathring{s}_{e_3} + \mathring{s}_{\not e_3}\right) + \left(\mathring{s}_{e_3} + \mathring{s}_{\not e_3}\right) \wedge e_3 = s \wedge e_3 \tag{8.46a}$$

$$s\mathring{s}_{e_3} + \mathring{s}_{\not e_3} e_3 = s e_3 \tag{8.46b}$$

$$s\left(\mathring{s} \cdot e_3\right) - \mathring{s}_{\not e_3} = -s \tag{8.46c}$$

$$\therefore s = \frac{\mathring{s}_{\not e_3}}{1 + \mathring{s} \cdot e_3} \tag{8.46d}$$

8.3.2 Network Parameter Conversion

Observe that the six cardinal poles of the Riemann sphere map to special points in the complex S-plane,

$$S = \begin{cases} +1 & \mathring{s} = e_1 \\ +j & \mathring{s} = e_2 \\ \infty & \mathring{s} = e_3 \\ -1 & \mathring{s} = -e_1 \\ -j & \mathring{s} = -e_2 \\ 0 & \mathring{s} = -e_3 \end{cases} \tag{8.47}$$

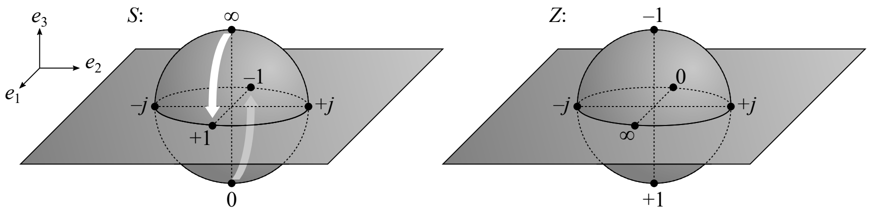


Figure 8.8 Identification of complex (a) reflection/scattering coefficient and (b) normalized impedance associated with the Cartesian poles of the Riemann sphere. The impedance is related to the scattering parameter by a 90-degree rotation, as shown by the white arrow.

If S is interpreted as a scattering parameter (reflection coefficient), then we may convert this to normalized impedance using (8.34),

$$Z = \begin{cases} \infty & \mathring{z} = e_1 \\ +j & \mathring{z} = e_2 \\ -1 & \mathring{z} = e_3 \\ 0 & \mathring{z} = -e_1 \\ -j & \mathring{z} = -e_2 \\ +1 & \mathring{z} = -e_3 \end{cases} \tag{8.48}$$

The poles are thus labeled in Figure 8.8. One may recognize a pattern in that the Z values are the same as the S values rotated about the e_2 axis by 90 degrees. We may therefore affect a conversion from s-parameters to z-parameters using the double-sided rotor transformation formula from Section 5.3,

$$\mathring{z} = e^{e_{13}\pi/4}\mathring{s}e^{-e_{13}\pi/4} \tag{8.49}$$

To verify this, let us write $\mathring{s}$ in terms of s $(= \alpha e_1 + \beta e_2)$ using (8.44),

$$\mathring{z} = e^{e_{13}\pi/4}\mathring{s}e^{-e_{13}\pi/4} = \tfrac{1}{2}\left(1 + e_{13}\right)\left(\tfrac{s^2+1}{s^2-1}e_3 - \tfrac{2}{s^2-1}s\right)\left(1 - e_{13}\right) \tag{8.50a}$$

$$= \tfrac{1}{2(s^2-1)}\left(1 + e_{13}\right)\left(\left(s^2+1\right)e_3 - 2s\right)\left(1 - e_{13}\right) \tag{8.50b}$$

$$= \tfrac{1}{(s^2-1)}\left(-\left(s^2+1\right)e_1 - s - e_{13}s + se_{13} + e_{13}se_{13}\right) \tag{8.50c}$$

$$= -\tfrac{1}{(s^2-1)}\left(\left(s^2+1\right)e_1 + 2\alpha e_3 + 2\beta e_2\right) \tag{8.50d}$$

Then we write z in terms of $\mathring{z}$ using (8.46d),

$$z = \frac{\left(\mathring{z} \wedge e_3\right) e_3^{-1}}{1 + \mathring{z} \cdot e_3} = \frac{\left(\tfrac{1}{s^2-1}\left(\left(s^2+1\right)e_1 + 2\alpha e_3 + 2\beta e_2\right) \wedge e_3\right) e_3}{1 - \tfrac{1}{s^2-1}\left(\left(s^2+1\right)e_1 + 2\alpha e_3 + 2\beta e_2\right) \cdot e_3} \tag{8.51a}$$

$$= \frac{-\left(s^2+1\right)e_1 - 2\beta e_2}{s^2 - 1 + 2\alpha} = \frac{\left(-\alpha^2 - \beta^2 + 1\right)e_1 + 2\beta e_2}{\alpha^2 + \beta^2 + 1 - 2\alpha} e^1 e_1 \tag{8.51b}$$

$$= \frac{-\alpha^2 - \beta^2 + 1 + 2\beta e_{12}}{\alpha^2 + \beta^2 + 1 - 2\alpha} e_1 = \frac{1 + S - S^* - S^2}{1 - S - S^* + S^2} e_1 \tag{8.51c}$$

Therefore,

$$Z = z e_1^{-1} = \frac{1 + S - S^* - S^2}{1 - S - S^* + S^2} = \frac{(1+S)\left(1-S^*\right)}{(1-S)\left(1-S^*\right)} \tag{8.52}$$

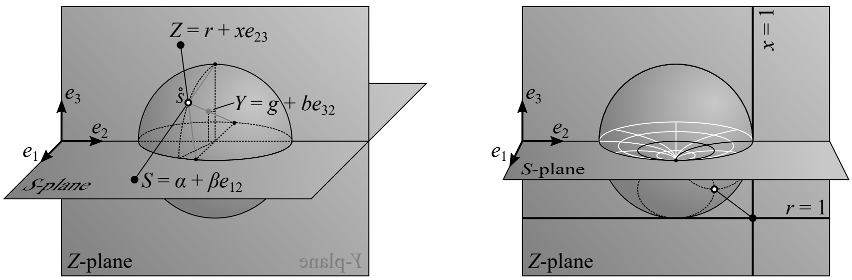
If we allow the term $(1 - S^*)$ that is common to the numerator and denominator to cancel, then we will have reproduced the classical load impedance formula (8.34). However, in the context of spacetime algebra, the denominator that remains $(1 - S)$ is no longer a scalar but a paravector having grade-0 and grade-2 parts, an object for which we have not yet clearly defined the operation of division. In principle, it would require deriving the square magnitude of the denominator by multiplying back in its complex conjugate $(1 - S)^*$ — precisely the terms we were about to cancel! Nevertheless, it seems apparent that the rotation in (8.49) effectively transforms scattering parameters into impedance.

While the above proof may seem rather messy, there is an easy way to shortcut the entire process. Knowing now that the Riemann spheres for z-parameters and s-parameters differ only by their orientation, we may skip the rotation entirely, and instead down-project the s-parameter Riemann sphere onto the plane normal to e_1 instead of to e_3, and with the reference orientation (the real part) given by $-e_3$ instead of e_1,

$$Z = z e_3 = \frac{\mathring{s}_{\not{e}_1}}{1 + \mathring{s} \cdot e_1} e_3 = r + x e_{23} \tag{8.53}$$

The imaginary unit of Z in this case is e_{23}.

It is well known that the scattering parameters for a normalized admittance differ by 180 degrees from the scattering parameters for an equivalent normalized impedance. We may thus deduce that admittance is obtained by down-projection



(a) (b)

Figure 8.9 (a) Recovery of s-parameters, z-parameters, and y-parameters from points on the Riemann sphere by stereographic down-projection onto different Cartesian planes. (b) Tracing $r = 1$ and $x = 1$ contours from the Z-plane onto the Riemann sphere and then into the S-plane, thus creating a Smith chart.

from the Riemann sphere onto the plane normal to $-e_1$, with reference orientation $-e_3$,

$$Y = ye_3 = \frac{\mathring{s}_{\cancel{e_1}}}{1 - \mathring{s} \cdot e_1} e_3 = g + be_{32} \tag{8.54}$$

We thus see that the Riemann sphere contains the s-parameters, z-parameters, and y-parameters directly, if merely interpreted against different complex planes; no rotation is required to convert them. This is illustrated in Figure 8.9.

8.3.3 Stereographic Null Vector

So far, we have only used three of the spacetime dimensions. It has been demonstrated that this is sufficient to perform the homogeneous Möbius transformations that convert reflection coefficients into impedance, or impedance to admittance, and so on, using only plane rotations [10]. The fourth dimension, having a positive square norm, is needed to give us access to inhomogeneous operations such as the additions of series impedance and shunt admittance to a known immittance point in the same manner. In the mathematical lexicon, we will have constructed a conformal geometric algebra (which is incidentally identical to spacetime algebra in this case) overlaid upon the base complex algebra of terminal parameters (scattering and immittance) that we wish to describe.

Therefore, in addition to stereographic projection onto the unit spatial Riemann sphere, let us introduce a constant offset into the temporal[4] dimension, e_0,

$$p = \mathring{s} + e_0 \tag{8.55}$$

For clarity, I will refer to p as the *spacetime Riemann parameter* or p-parameter, to distinguish it from the immittance and scattering parameters. The spacetime Riemann parameter contains the same information as immittance and scattering parameters, but in a form that is more easily manipulated. Note that since $\mathring{s}$ has unit magnitude with negative signature, p is always a null vector,

$$p^2 = (\mathring{s} + e_0)^2 = \mathring{s}^2 + e_0^2 = -1 + 1 = 0 \tag{8.56}$$

To recover the s-, z-, or y-parameters from the p-parameter, we may use modified forms of (8.46d), (8.53), and (8.54) that additionally reject and normalize to the temporal axis,

$$S = \frac{\mathring{s}_{\cancel{e_3}}}{1 + \mathring{s} \cdot e_3} e^1 = \frac{p_{\cancel{e_{30}}}}{p \cdot (e_0 + e_3)} e^1 \tag{8.57a}$$

$$Z = \frac{\mathring{s}_{\cancel{e_1}}}{1 + \mathring{s} \cdot e_1} e_3 = \frac{p_{\cancel{e_{10}}}}{p \cdot (e_0 + e_1)} e_3 \tag{8.57b}$$

$$Y = \frac{\mathring{s}_{\cancel{e_1}}}{1 - \mathring{s} \cdot e_1} e_3 = \frac{p_{\cancel{e_{10}}}}{p \cdot (e_0 - e_1)} e_3 \tag{8.57c}$$

The full set of conversions between the p-parameter and conventional immittance or scattering parameters is given in Table 8.2.

8.4 ROTORS FOR NETWORK ANALYSIS

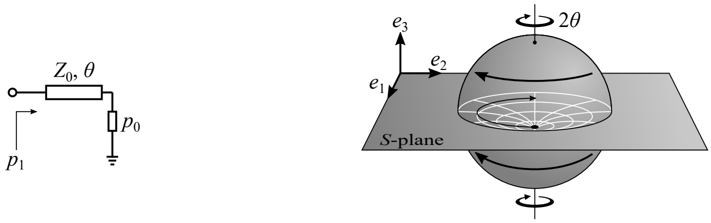
Having expended the effort to define a completely new representation of terminal behavior in microwave networks, the p-parameters, we are now prepared to fulfill the promise that this representation would allow easy implementation of operations such as cascading lumped elements, transmission lines, and stubs. Each of these operations acts upon a known p-point (effectively a known immittance or reflection coefficient) as a simple rotation in the four-dimensional space we have constructed. Such rotors are more generally called *versors* in conformal geometric algebra.

4 Keep in mind that this is merely an application by analogy of spacetime algebra. There is actually nothing at all temporal about the fourth dimension here, except that it occupies the slot within the metric signature having opposite sign to the other spatial dimensions.

Table 8.2

Conversions Between Spacetime Riemann Parameters and Immittance/Scattering Parameters

Immittance/Scattering Parameter		Spacetime Riemann Parameter	
Complex Form	Vector Form	Conversion to p	Conversion from p
$S = \alpha + \beta e_{12}$	$s = \alpha e_1 + \beta e_2$	$p = \frac{s^2+1}{s^2-1} e_3 - \frac{2s}{s^2-1} + e_0$	$S = \frac{p_{e_{30}}}{p\cdot(e_0+e_3)} e^1$
$Z = r + x e_{23}$	$z = -r e_3 + x e_2$	$p = \frac{z^2+1}{z^2-1} e_1 - \frac{2z}{z^2-1} + e_0$	$Z = \frac{p_{e_{10}}}{p\cdot(e_0+e_1)} e_3$
$Y = g + b e_{32}$	$y = -g e_3 - b e_2$	$p = \frac{1+y^2}{1-y^2} e_1 + \frac{2y}{1-y^2} + e_0$	$Y = \frac{p_{e_{10}}}{p\cdot(e_0-e_1)} e_3$



(a) (b)

Figure 8.10 (a) Schematic of a matched transmission line cascaded with a known immittance having p-parameter p_0. (b) The resulting phase delay is modeled by a rotation in the S-plane by an angle twice the electrical length of the transmission line.

8.4.1 Phase Delay

The simplest case we will explore first as a warm-up for what is to follow is the phase rotation introduced by cascading a matched transmission line, as shown in Figure 8.10(a). In the S-plane, this manifests as a rotation around the origin through an angle that is twice the electrical length of the transmission line (i.e., the round-trip phase of the reflection). This is illustrated in Figure 8.10(b), and is implemented mathematically as a rotation using θe_{21} as the generator,

$$T_0 = \theta e_{21} \tag{8.58}$$

Thus, the p-point seen at the input of the cascaded transmission line is

$$p_1 = e^{T_0} p_0 e^{-T_0} = e^{\theta e_{21}} p_0 e^{-\theta e_{21}} \tag{8.59}$$

8.4.2 Series Impedance

The true benefit of a conformal geometric algebra is that some operations which do not conform to simple rotations in the base space (as the phase delay did in Section 8.4.1) are nevertheless simple rotations in the higher dimensional space. Take, for example, the addition of a normalized series resistance, r. This amounts to a translation in the Z-plane through a distance r along the $-e_3$ direction. Addressed to points on the spacetime Riemann sphere, it can be shown that the generator for this rotation is

$$R = -\tfrac{r}{2} e_3 \left(e_0 + e_1\right) \tag{8.60}$$

One could validate this result by first projecting an initial impedance, $z = r_0 + jx_0$, onto the spacetime Riemann sphere, performing the rotation described above, and then projecting back down to the complex z plane. This undoubtedly tedious exercise is left to the reader.

What is more important is to recognize the pattern revealed in the elements of the generator,

- r is the translation distance.
- $-e_3$ is the translation direction.
- $e_0 + e_1$ is a null vector given by the sum of the temporal axis, e_0, and the reference normal, or Riemann projection pole, e_1, to the plane in which translation occurs (in this case, the Z-plane).

This allows us to infer the generator for the addition of a series reactance, or a translation in the Z-plane by a distance x along the e_2 direction,

$$X = \tfrac{x}{2} e_2 \left(e_0 + e_1\right) \tag{8.61}$$

One might wonder whether we can define a single generator for a complex series impedance, amounting to the addition of a series resistance and a series reactance at the same time. This simply requires compounding the two rotors as one,

$$R_{r+jx} = e^R e^X \tag{8.62}$$

Recall that the product-of-powers rule applies only if the two exponentiated arguments commute. In fact, the generators R and X do commute trivially as they both contain the same null vector, $e_0 + e_1$,

$$RX = -\tfrac{r}{2}e_3\left(e_0 + e_1\right)\tfrac{x}{2}e_2\left(e_0 + e_1\right) = \tfrac{rx}{4}e_3e_2\left(e_0 + e_1\right)^2 = 0 \tag{8.63}$$

Therefore, $RX = XR = 0$, the product-of-powers rule applies, and the combined generator may be written

$$R + X = \left(\tfrac{x}{2}e_2 - \tfrac{r}{2}e_3\right)\left(e_0 + e_1\right) \tag{8.64}$$

This confirms our expectation that the order of the addition, resistance first or reactance first, should not matter.

8.4.3 Parallel Admittance

We can easily apply rotations that have the effect of adding shunt conductance and susceptance, which are translations in the Y-plane (reference normal $-e_1$) by distances g and b in the $-e_3$ and $-e_2$ directions, respectively. Based on the pattern established above, these are given by the rotation generators below,

$$G = -\tfrac{g}{2}e_3\left(e_0 - e_1\right) \tag{8.65a}$$

$$B = -\tfrac{b}{2}e_2\left(e_0 - e_1\right) \tag{8.65b}$$

Once again, $e_0 - e_1$ is a null vector, and as it is common to both generators, these two operations trivially commute.

8.4.4 Carter Rotations

Presumably, the reader familiar with microwave engineering has understood that rotations associated with the generators R, X, G, and B would trace out the familiar contours of constant reactance, resistance, susceptance, and conductance, respectively, of the Smith chart in the S-plane, as shown in Figure 8.11(a, b).

For completeness, there is one other plane (having two reference directions), representing two degrees of freedom not yet explored by the above rotations. Generators associated with these two remaining degrees of freedom may be written as follows,

$$Q = \tfrac{q}{2}e_{23} \tag{8.66a}$$

$$N = \tfrac{\ln(n)}{2}e_{10} \tag{8.66b}$$

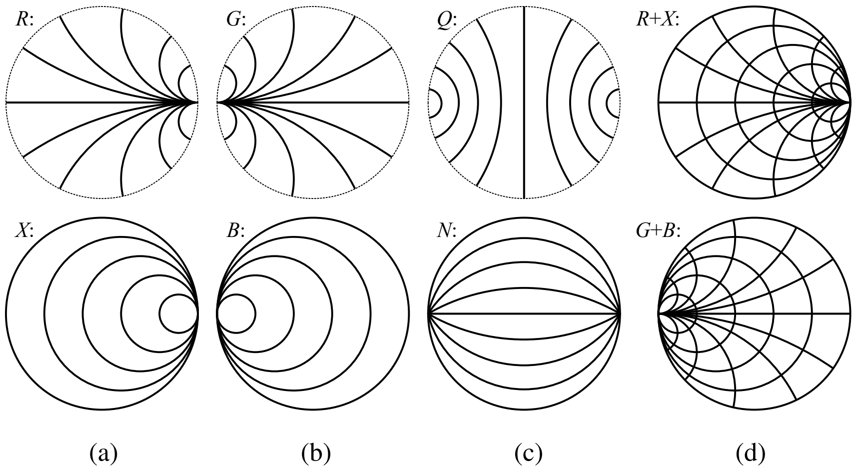


Figure 8.11 Contours in the S-plane traced out by (a) impedance rotations (generators R and X), (b) admittance rotations (generators G and B), and (c) Carter rotations (generators Q and N). (d) Smith impedance chart (top) formed by combining contours from generators R and X, and Smith admittance chart (bottom) formed by combining contours from generators G and B.

The first generator, Q, traces out contours of constant impedance magnitude, $|Z|$, while effecting an increase in its phase,

$$Z \to Ze^{jq} \tag{8.67}$$

The second generator, N, traces out contours of constant impedance phase, while effecting an increase in its magnitude,

$$Z \to nZ \tag{8.68}$$

The contours associated with these generators are shown in Figure 8.11(c). Although less familiar, they are the contours of a modification to the Smith chart known as a *Carter chart*, which is used to assist in developing optimal broadband impedance-matching networks [4, 11]. For simplicity, I shall refer to these operators as *Carter rotations*.

Unlike the previous cases, the product of these two generators is not null, but they do nevertheless commute with one another ($QN = NQ$). Thus, in the same way that series resistance and reactance rotations (R and X) can be applied in any order, successive Carter rotations may combine in any order with identical results.

The commuting pairs of generators, combined into a single plot, create useful and familiar nomograms for working with immittances, such as the familiar Smith impedance and admittance charts shown in Figure 8.11(d). My intention is not to teach the use of the Smith charts in detail (for that, the reader is referred to any of numerous textbooks on the subject [4–7]), only to illustrate that they can be derived from the mathematical ideas and principles described in this book.

8.4.5 Properties of Smith and Carter Rotations

The rotation generators R, X, G, B, Q, and N form a closed bivector group [10] that obeys a number of useful properties (for the purposes of this section, we assume normalized generator magnitudes, i.e., $r = x = g = b = q = \ln(n) = 2$). First, as has already been implied, they can be arranged into orthogonal pairs,

$$R \cdot X = G \cdot B = Q \cdot N = 0 \tag{8.69}$$

which is sufficient to guarantee that successive transformations from any pair commute. They are also related by duality,

$$R = Xi \tag{8.70a}$$

$$G = Bi \tag{8.70b}$$

$$N = Qi \tag{8.70c}$$

where

$$i = e_{0123} \tag{8.71}$$

Another set of relationships is rooted in the process of relative reversion (see Section D.5.6),

$$G = R^\dagger \tag{8.72a}$$

$$B = -X^\dagger \tag{8.72b}$$

$$N = N^\dagger \tag{8.72c}$$

$$Q = -Q^\dagger \tag{8.72d}$$

Finally, all of these generators are simple bivectors, with lightlike, timelike, and spacelike scalar norms as follows,

$$R^2 = X^2 = G^2 = B^2 = 0 \tag{8.73a}$$

$$N^2 = -Q^2 = 1 \tag{8.73b}$$

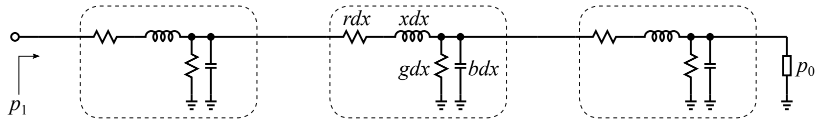


Figure 8.12 Distributed-element transmission-line model.

8.4.6 Unmatched Cascaded Transmission Line

We have thus far identified rotations applicable to series and shunt connections of lumped elements, and we began this section with a simple phase rotation for a matched transmission line in cascade. We may now use those results to derive a rotation describing the effect of an unmatched cascaded transmission line.

To that end, we consider the distributed lumped-element model for a transmission line shown in Figure 8.12. In this case, the scalar immittance quantities r, x, g, and b are interpreted as immittance densities, or immittance per unit length. We assume a total line length of l, and take the limit as the number of unit cells approaches infinity (i.e., the discrete immittance elements become infinitesimal).

Consider the rotor for a single unit cell,

$$e^{dU} = e^{dR}e^{dX}e^{dG}e^{dB} = e^{dR+dX}e^{dG+dB} \tag{8.74}$$

Note that we have applied the product-of-powers rule to the impedance and admittance generator pairs, since we know that those terms commute. However, the admittance pair does not in general commute with the impedance pair, so we have kept those separate for now. This would seem to be a roadblock to making further progress, but it turns out that even these terms commute in the limit in which they become very small, as is the case here.

To see this, note that the transformation of a vector p by a small rotation dA may be expanded as follows [10, 12],

$$e^{dA}pe^{-dA} \approx (1 + dA)p(1 - dA) \tag{8.75a}$$

$$= p + (dAp - pdA) - dApdA \approx p + 2\left[dA, p\right] \tag{8.75b}$$

where the square brackets denote the commutator product and we have neglected the quadratic term ($dApdA$) for small dA. Thus, for two successive small rotations, dA and dB, we have

$$e^{dB}e^{dA}pe^{-dA}e^{-dB} \approx e^{dB}\left(p + 2\left[dA, p\right]\right)e^{-dB} \tag{8.76a}$$

$$\approx (p + 2\,[dA, p]) + 2\,[dB, p + 2\,[dA, p]] \tag{8.76b}$$

$$= p + 2\,[dA, p] + dB\,(p + 2\,[dA, p]) - (p + 2\,[dA, p])\,dB \tag{8.76c}$$

$$= p + 2\,[dA, p] + 2\,[dB, p] + dB\,(2\,[dA, p]) - (2\,[dA, p])\,dB \tag{8.76d}$$

$$\approx p + 2\,[dA + dB, p] = e^{dA+dB} p e^{-dA-dB} \tag{8.76e}$$

showing that commutation applies for the dominant, lowest-order terms.

We may thus write the combined generator for the distributed-element model of a length-l transmission line quite simply,

$$T = l dU = l(dR + dX + dG + dB) \tag{8.77a}$$

$$= \tfrac{l}{2}\,(xe_2 - re_3)\,(e_0 + e_1) - \tfrac{l}{2}\,(be_2 + ge_3)\,(e_0 - e_1) \tag{8.77b}$$

$$= \tfrac{l}{2}(x - b)e_{20} + \tfrac{l}{2}(x + b)e_{21} + \tfrac{l}{2}(r - g)e_{13} + \tfrac{l}{2}(r + g)e_{03} \tag{8.77c}$$

$$= \tfrac{l}{2}(x - b)e_{20} + \tfrac{l}{2}(x + b)e_{21} + \tfrac{l}{2}(r - g)ie_{20} + \tfrac{l}{2}(r + g)ie_{21} \tag{8.77d}$$

$$= \tfrac{l}{2}\,[(x + ir) - (b + ig)]\,e_{20} + \tfrac{l}{2}\,[(x + ir) + (b + ig)]\,e_{21} \tag{8.77e}$$

$$= \tfrac{\beta l}{2}\,(z_0 - y_0)\,e_{20} + \tfrac{\beta l}{2}\,(z_0 + y_0)\,e_{21} \tag{8.77f}$$

where β, z_0, and y_0 are the phase constant, the normalized characteristic impedance, and the normalized characteristic admittance of the transmission line, respectively, given by

$$\beta = \sqrt{(x + ir)(b + ig)} \tag{8.78a}$$

$$z_0 = y_0^{-1} = \sqrt{\frac{x + ir}{b + ig}} \tag{8.78b}$$

(We have allowed division here since the pseudoscalar i commutes like a scalar with all the relevant terms.) These equations match those we expect from classical transmission-line theory [4, 5] under the assumption $i = -j$ (which is valid since all we require is that $i^2 = j^2 = -1$). Contours in the S-plane for this rotation under various typical conditions are given in Figure 8.13.

Note that when $r = g = 0$ (the line is lossless) and $x = b$ (the line is matched), we have

$$T = \tfrac{l}{2} x e_{21} = \tfrac{\beta l}{2} e_{21} = \theta e_{21} = T_0 \tag{8.79}$$

matching the simple phase delay generator we derived in Section 8.4.1. Rotation generators for the Smith, Carter, and transmission-line rotations are summarized in Table 8.3.

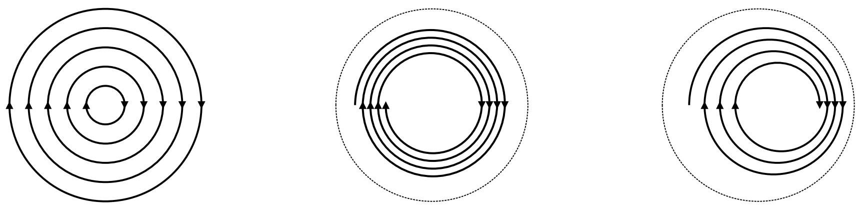


Figure 8.13 Contours in the S-plane traced out by the T rotation generator for increasing l, when (a) $r = g = 0$ and $x = b$, corresponding to a matched, lossless transmission line, (b) r and g are small while $x = b$ corresponding to a matched, slightly lossy line, and (c) r and g are small while $x > b$, corresponding to a high-impedance, slightly lossy line.

Table 8.3

Rotation Generators for Network Analysis

Description	Complex Operation	Rotation Generator
Series resistance	$Z + r$	$R = \frac{r}{2}\left(e_0 + e_1\right) e_3$
Series reactance	$Z + jx$	$X = \frac{x}{2} e_2 \left(e_0 + e_1\right)$
Parallel conductance	$Y + g$	$G = \frac{g}{2}\left(e_0 - e_1\right) e_3$
Parallel susceptance	$Y + jb$	$B = \frac{b}{2}\left(e_0 - e_1\right) e_2$
Impedance magnitude	nZ	$N = \frac{\ln(n)}{2} e_{10}$
Impedance phase	Ze^{jq}	$Q = \frac{q}{2} e_{23}$
Cascade transmission line	$z_0 \frac{Z + jz_0 \tan(\beta l)}{z_0 + jZ \tan(\beta l)}$	$T = \frac{\beta l}{2}\left(z_0 - y_0\right) e_{20} + \frac{\beta l}{2}\left(z_0 + y_0\right) e_{21}$
	$= \frac{l}{2}(x - b)e_{20} + \frac{l}{2}(x + b)e_{21} + \frac{l}{2}(r - g)e_{13} + \frac{l}{2}(r + g)e_{03}$	

8.4.7 Nonsimple Bivector Generators

The transmission-line generator, T, is unique among all the generators we have described in this chapter in that it is the only one that is generally a nonsimple bivector. Thus, while the rest can be expanded into a rotor using Euler's formula in either its purely trigonometric or hyperbolic form, depending on the signature of the square norm (the first shown explicitly in Section 5.1.4), one must exercise some care in expanding the rotor for T.

Consider an arbitrary bivector generator, Q, and its polar canonical form,

$$Q = \mathbf{q}e^{i\varphi} = \mathbf{q}z \tag{8.80}$$

where $z = e^{i\varphi} = \cos\varphi + i\sin\varphi$. The manner in which we defined the canonical form in Section 5.1.7 ensures not only that $\mathbf{q}$ is simple, but that its scalar square norm is positive. Let us further factor out the magnitude of this norm as a separate scalar, θ, such that

$$Q = \hat{\mathbf{q}}\theta z \tag{8.81}$$

where $\theta > 0$ and $\hat{\mathbf{q}}^2 = 1$. We may now apply this to the Taylor expansion of the rotor,

$$e^Q = \sum_{n=0}^{\infty} \frac{Q^n}{n!} = 1 + Q + \frac{Q^2}{2!} + \frac{Q^3}{3!} + \frac{Q^4}{4!} + \cdots \tag{8.82a}$$

$$= \left(1 + \frac{Q^2}{2!} + \frac{Q^4}{4!} + \cdots\right) + \left(Q + \frac{Q^3}{3!} + \frac{Q^5}{5!} + \cdots\right) \tag{8.82b}$$

$$= \left(1 + \frac{(\theta z)^2}{2!} + \frac{(\theta z)^4}{4!} + \cdots\right) + \hat{\mathbf{q}}\left((\theta z) + \frac{(\theta z)^3}{3!} + \frac{(\theta z)^5}{5!} + \cdots\right) \tag{8.82c}$$

$$= \cosh(\theta z) + \hat{\mathbf{q}}\sinh(\theta z) \tag{8.82d}$$

where (for computer codes) the functions $\cosh$ and $\sinh$ must accept the fully complex argument, θz.

The nonsimple nature of T is what allows it to encapsulate both rotation and dilation independently within the same rotor, leading to the spiral, or *loxodromic* contours in Figures 8.13(b, c). All other rotors described previously produced only circular contours.

While the rotor-based technique for calculating immittances of complex networks presented in this section is unlikely to ever entirely replace conventional methods [13], there will inevitably be cases for which it is especially well-suited, if we may only be prepared to recognize them. One application which has already been met with some success is a novel method based on conformal geometric algebra for the interpolation of discrete measured data which is both more accurate and more efficient than traditional numerical techniques [14, 15].

Perhaps more importantly, the approach outlined here may serve as an example of the power of geometric algebra to illuminate complex mathematics in intuitive ways. Special relativity has revealed a higher-dimensional realm in which the laws of electromagnetics more naturally reside. As we adapt our minds to visualizing those higher dimensions in order to improve our understanding, so too should our mathematical vocabulary expand to describe higher-dimensional objects, whether physical or abstract, in a natural and consistent way.

References

[1] J. Dressel, K. Y. Bliokh, and F. Nori, "Spacetime algebra as a powerful tool for electromagnetism," *Physics Reports*, vol. 589, pp. 1–71, 2015.

[2] D. Hestenes, "Multivector calculus," *Journal of Mathematical Analysis and Applications*, vol. 24, no. 2, pp. 313–325, 1968.

[3] D. Hestenes, "Differential forms in geometric calculus," *Proceedings of the Third Conference on Clifford Algebras and their Applications in Mathematical Physics*, vol. 55, pp. 269–285, 1993.

[4] M. A. Morgan, *Principles of RF and Microwave Design*. Norwood, MA: Artech House, November 2019.

[5] D. M. Pozar, *Microwave Engineering*, 4th ed. New York: Wiley, 2011.

[6] S. Ramo, J. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics*. New York: Wiley, 1984.

[7] G. Gonzalez, *Microwave Transistor Amplifiers: Analysis and Design*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1996.

[8] Wikipedia. (2024, January) Möbius transformation. https://en.wikipedia.org/wiki/M%C3%B6bius_transformation.

[9] M. A. Morgan, *Reflectionless Filters*. Norwood, MA: Artech House, 2017.

[10] A. Arsenovic, "Applications of conformal geometric algebra to transmission line theory," *IEEE Access*, vol. 5, pp. 19 920–19 941, July 2017.

[11] Engineering and Technology History Wiki. (2015) History of broadband impedance matching. http://ethw.org/History_of_Broadband_Impedance_Matching.

[12] L. Dorst, D. Fontijne, and S. Mann, *Geometric Algebra for Computer Science*. Burlington, MA: Morgan Kaufmann, 2007.

[13] F. F. Kuo, *Network Analysis and Synthesis*, 2nd ed. New York: Wiley, 1962.

[14] A. Arsenovic, "A new method for interpolating transmissive two-port networks," *810 Labs, LLC*, May 2018.

[15] A. Arsenovic, "A method for interpolating transmissive, two-port networks," U.S. Provisional Patent Application 62/663,575, April 27, 2018.

Epilogue

I feel compelled at this stage to make some concluding remarks about the impact I hope this book will have on those who may one day read it. Undoubtedly there will be those who scoff at the very notion that relativity could have any practical value in the field of engineering, assuming that any time spent on it would, for them, be wasted. Others may hold the same skepticism while taking a kinder view of the spirit of exploration that drives some of us to expand our knowledge, however useless it may be. I urge you not to fall into either category.

What this book does is provide a model of electromagnetic phenomenon that is new to most practicing engineers, but is nevertheless completely compatible with the classical view we are taught in college. What value is it, then? What can it add to a theory with which it is entirely equivocal? What it adds is perspective.

The value of that perspective should not be underestimated. We tend to assume, given our Newtonian mindset, that relativistic speeds are unapproachable at our current level of technology, and that as a result relativistic effects will never be significant. Hopefully, having read Section 6.2.4 on the Fizeau water experiment, and Section 3.3 on satellite navigation systems, the reader will now understand how wrong that assumption is. Light moves at the speed of light, and as electrical engineers, light is our business. Light itself is subject to relativistic effects which are not only measurable, but (I believe) exploitable, if only we can adapt our way of thinking to a more relativistic one.

As I have said, the relativistic formulation of Maxwell's equation(s) developed in this book contradicts nothing described by the classical Gibbs-Heaviside formulation. On the contrary, it affirms that formulation, even when it goes against our natural-born instincts. But, as physicists in the early twentieth century discovered, when relativity is not brought out explicitly, when electromagnetic theory is constructed on space-in-time instead of spacetime, the temptation is strong to disregard the complex field calculations that formulation requires, and instead to take a mental shortcut through Newtonian assumptions about wave propagation. That is essentially what the engineers who built the original GPS system were doing when they conceived of their system in purely geometric terms, forcing them to correct it later with post-Newtonian offsets. In doing so, they missed the opportunity to create an even more accurate and enduring system, one that could one day enable applications beyond even their fertile imaginations.

So, without introducing any new physics, what this book offers instead is a new way of looking at the familiar phenomenon of electromagnetics. It is unquestionably an elegant viewpoint, arguably a convenient one, and I think potentially a very useful one. But is it real? No doubt, many who read this will continue to struggle to put themselves mentally in the relativistic world, and will wonder deep down whether it is the way things truly are, or else merely a fantastical model that happens to be compatible with natural laws. The question of reality is a metaphysical one, which I could easily dismiss as beyond the scope of this book. Perhaps so, but I will offer my answer anyway,

> *At any given time, the model that is most real is the one that explains all known phenomena within its scope in the most elegant form.*

Why should one accept this definition? What value is elegance in a scientific theory? A scientist might call it a corollary of Occam's razor. A man of faith might say that the subjective quality of elegance is nothing less than the holy spirit revealing to us that which is right and true in a universe purpose-built to please its creator. As I am both a scientist and a man of faith, I consider these answers one and the same.

Therefore, contrary to our corporeal perceptions, we do not live in a three-dimensional universe evolving in time (the paravector view), rather we exist in a four-dimensional spacetime. All that we see and hear and experience are merely projections from the higher-dimensional realm into our three-dimensional awareness. The electric and magnetic fields are in fact the projections (literally, using mathematical projection formula) of a parent construct known as the Faraday bivector field, a directed area in spacetime, onto our particular worldline. They are like two facets of the same jewel.

Appendix A

Gaussian-CGS Units

From the perspective of one familiar with SI units, the so-called *Gaussian centimeter-gram-second (CGS)* unit system is more than simply a different set of physical units. It is a redefinition of certain fundamental physical quantities and their associated formulas, in particular the relationship between electromagnetic units or phenomena and mechanical units or phenomena.

Aside from the relatively straightforward substitution of centimeters for meters, and grams for kilograms, the Gaussian units also absorb factors of 4π, which are common in electromagnetic calculations, into certain quantities, thus removing them from some formulas (and adding them to others). The vacuum permittivity, ε_0, and permeability, μ_0, are also absorbed into the definitions of the charge and field quantities such that these constants no longer appear in any expressions (but the vacuum speed of light, c, does appear frequently) [1].

I find it useful to think of the Gaussian system as simply a relabeling of physical quantities with scaled versions that incorporate certain fundamental constants. One could, for example, imagine a system in which length is not measured in meters or inches, but in the time it takes a beam of light to traverse it in a vacuum. In this hypothetical system, then the new length in seconds, l', would be related to the original as $l' = l/c$. Since c is fundamental and fixed, there is no ambiguity in the distance that is being described, however strange it may seem.

In Tables A.1 to A.3, primed variables represent the Gaussian-CGS definitions, while those unprimed represent the more familiar SI definitions. Note that although different Gaussian unit names are conventionally used to distinguish them, the electromagnetic fields and flux densities ($\mathbf{E}'$, $\mathbf{H}'$, $\mathbf{D}'$, and $\mathbf{B}'$) are all expressed in equivalent units — that is, 1 statV/cm = 1 Fr/cm^2 = 1 Oe = 1 G = 1 cm$^{-1/2}$g$^{1/2}$s^{-1}.

Table A.1

Definition of Classical Gaussian Quantities in Terms of SI Quantities [2]

Quantity	Gaussian Definition	Gaussian Unit
Electric charge	$q' = \frac{q}{\sqrt{4\pi\varepsilon_0}}$	Fr (franklin) or statC (statcoulomb)
Current	$i' = \frac{i}{\sqrt{4\pi\varepsilon_0}}$	Fr/s
Scalar electric potential	$\varphi' = \varphi\sqrt{4\pi\varepsilon_0}$	statV (statvolt)
Vector magnetic potential	$\mathbf{A}' = \mathbf{A}\sqrt{\frac{4\pi}{\mu_0}}$	G·cm
Voltage	$v' = v\sqrt{4\pi\varepsilon_0}$	statV
Electric field	$\mathbf{E}' = \mathbf{E}\sqrt{4\pi\varepsilon_0}$	statV/cm
Electric flux density/displacement	$\mathbf{D}' = \mathbf{D}\sqrt{\frac{4\pi}{\varepsilon_0}}$	Fr/cm^2
Magnetic field	$\mathbf{H}' = \mathbf{H}\sqrt{4\pi\mu_0}$	Oe (oersted)
Magnetic flux density	$\mathbf{B}' = \mathbf{B}\sqrt{\frac{4\pi}{\mu_0}}$	G (gauss)
Magnetic flux	$\Phi' = \Phi\sqrt{\frac{4\pi}{\mu_0}}$	Wb (webers)
Resistance	$R' = 4\pi\varepsilon_0 R$	s/cm
Capacitance	$C' = \frac{C}{4\pi\varepsilon_0}$	cm
Inductance	$L' = 4\pi\varepsilon_0 L$	s^2/cm
Electric susceptibility†	$\chi_e' = \frac{1}{4\pi}\chi_e$	(unitless)
Magnetic susceptibility†	$\chi_m' = \frac{1}{4\pi}\chi_m$	(unitless)

†Note that the electric and magnetic susceptibilities, though unitless in both systems, have values that differ by a factor of 4π.

Table A.2

Constitutive Relations in SI and Gaussian Units

SI Definition	Gaussian Definition
$\mathbf{D} = \varepsilon \mathbf{E}$	$\mathbf{D}' = \varepsilon \mathbf{E}'$
$= \varepsilon_0 \varepsilon_r \mathbf{E}$	$= \varepsilon_r \mathbf{E}'$
$= \varepsilon_0 (1 + \chi_e) \mathbf{E}$	$= (1 + 4\pi\chi_e') \mathbf{E}'$
$= (\varepsilon' - j\varepsilon'') \mathbf{E}^\dagger$	$= (\varepsilon' - j\varepsilon'') \mathbf{E}'$
$\mathbf{B} = \mu \mathbf{H}$	$\mathbf{B}' = \mu \mathbf{H}'$
$= \mu_0 \mu_r \mathbf{H}$	$= \mu_r \mathbf{H}'$
$= \mu_0 (1 + \chi_m) \mathbf{H}$	$= \left(1 + 4\pi\chi_M'\right) \mathbf{H}'$
$= (\mu' - j\mu'') \mathbf{H}$	$= (\mu' - j\mu'') \mathbf{H}'$

†Time-harmonic solutions are assumed when complex values are given.

Table A.3

Key Classical Equations in SI and Gaussian Units

Name	SI Equation	Gaussian Equation
Coulomb force	$F = \frac{q_1 q_2}{4\pi\varepsilon_0 r^2}$	$F = \frac{q_1' q_2'}{r^2}$
Lorentz force	$F = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$	$F = q'(\mathbf{E}' + \frac{1}{c}\mathbf{v} \times \mathbf{B}')$
Gauss's law	$\boldsymbol{\nabla} \cdot \mathbf{D} = \rho$	$\boldsymbol{\nabla} \cdot \mathbf{D}' = 4\pi\rho'$
Magnetic nondivergence	$\boldsymbol{\nabla} \cdot \mathbf{B} = 0$	$\boldsymbol{\nabla} \cdot \mathbf{B}' = 0$
Faraday's law	$\boldsymbol{\nabla} \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$	$\boldsymbol{\nabla} \times \mathbf{E}' = -\frac{1}{c}\frac{\partial \mathbf{B}'}{\partial t}$
Ampère's law	$\boldsymbol{\nabla} \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}$	$\boldsymbol{\nabla} \times \mathbf{H}' = \frac{4\pi}{c}\mathbf{J}' + \frac{1}{c}\frac{\partial \mathbf{D}'}{\partial t}$
Biot-Savart law	$\mathbf{B} = \frac{\mu_0}{4\pi}\oint \frac{I d\mathbf{l}' \times \mathbf{r}}{r^2}$	$\mathbf{B}' = \frac{1}{c}\oint \frac{I d\mathbf{l}' \times \mathbf{r}}{r^2}$
Poynting vector	$\mathbf{S} = \mathbf{E} \times \mathbf{H}$	$\mathbf{S}' = \frac{c}{4\pi}\mathbf{E}' \times \mathbf{H}'$
Electric potential	$\mathbf{E} = -\boldsymbol{\nabla}\varphi - \frac{\partial \mathbf{A}}{\partial t}$	$\mathbf{E}' = -\boldsymbol{\nabla}\varphi' - \frac{1}{c}\frac{\partial \mathbf{A}'}{\partial t}$

References

[1] M. Schwartz, *Principles of Electrodynamics*. New York: Dover, 1987.

[2] Wikipedia. (2024, January) Gaussian units. https://en.wikipedia.org/wiki/Gaussian_units.

Appendix B

Vector Calculus

B.1 VECTOR PRODUCTS

B.1.1 Dot Product

For any two vectors, $\mathbf{a}$ and $\mathbf{b}$, in three-dimensional space, the dot product (or *scalar product* or inner product) [1] is given by

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta \tag{B.1}$$

where θ is the Euclidean angle between the two vectors. In Cartesian coordinates,

$$\mathbf{a} \cdot \mathbf{b} = a_x b_x + a_y b_y + a_z c_z \tag{B.2}$$

B.1.2 Cross Product

The cross product or *vector product* is

$$\mathbf{a} \times \mathbf{b} = (|\mathbf{a}||\mathbf{b}| \sin\theta)\, \mathbf{c} \tag{B.3}$$

where $\mathbf{c}$ is the unit vector perpendicular to both $\mathbf{a}$ and $\mathbf{b}$ with orientation given by the right-hand rule (that is, in the direction indicated by the thumb of the right hand when the fingers are curled from $\mathbf{a}$ toward $\mathbf{b}$) when a right-handed coordinate system is used (which is most common) and by the corresponding left-hand rule when left-handed coordinates are used. In Cartesian coordinates,

$$\mathbf{a} \times \mathbf{b} = \mathbf{x}\left(a_y b_z - a_z b_y\right) + \mathbf{y}\left(a_z b_x - a_x b_z\right) + \mathbf{z}\left(a_x b_y - a_y b_x\right) \tag{B.4}$$

One may write the cross product in Cartesian coordinates by first forming the skew-symmetric matrix

$$\left[\mathbf{a}\right]_\times = -\left[\mathbf{a}\right]_\times^T = \begin{pmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{pmatrix} \tag{B.5}$$

Then,

$$\mathbf{a} \times \mathbf{b} = \left[\mathbf{a}\right]_\times \mathbf{b} = \begin{pmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{pmatrix} \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} = \begin{pmatrix} a_y b_z - a_z b_y \\ a_z b_x - a_x b_z \\ a_x b_y - a_y b_x \end{pmatrix} \tag{B.6}$$

Note also that

$$\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a} = -\left[\mathbf{b}\right]_\times \mathbf{a} = \left[\mathbf{b}\right]_\times^T \mathbf{a} = \left(\mathbf{a}^T \left[\mathbf{b}\right]_\times\right)^T \tag{B.7}$$

Alternatively, the ith component of the cross product can be written in tensor index notation with the help of the Levi-Civita symbol (see Appendix C),

$$\left(\mathbf{a} \times \mathbf{b}\right)^i = \epsilon_{ijk} a^j b^k \tag{B.8}$$

where $\epsilon_{123} = 1$.

B.1.3 Triple Products

The *scalar triple product* of vectors **a**, **b**, and **c** is

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) = \epsilon_{ijk} a^i b^j c^k \tag{B.9}$$

whereas the *vector triple product* is

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b}) \tag{B.10}$$

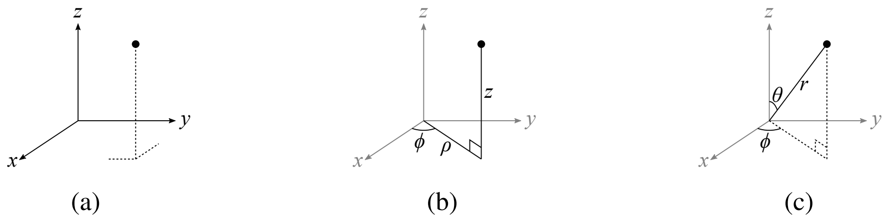


Figure B.1 Definition of coordinate systems used for vector calculus in this book. (a) Cartesian. (b) Cylindrical. (c) Spherical.

B.1.4 Dyadic Product

Finally, the dyadic product or outer product is

$$\mathbf{a} \otimes \mathbf{b} = \mathbf{a}\mathbf{b}^T = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} \begin{pmatrix} b_x & b_y & b_z \end{pmatrix} = \begin{pmatrix} a_x b_x & a_x b_y & a_x b_z \\ a_y b_x & a_y b_y & a_y b_z \\ a_z b_x & a_z b_y & a_z b_z \end{pmatrix} \quad \text{(B.11)}$$

having as its result a second-rank tensor. The commutation of two dyadic products in three dimensions gives the skew-symmetric matrix expansion of the cross product,

$$\mathbf{b} \otimes \mathbf{a} - \mathbf{a} \otimes \mathbf{b} = [\mathbf{a} \times \mathbf{b}]_\times \quad \text{(B.12)}$$

B.2 DIFFERENTIAL FORMS

Most spatial derivatives can be represented using the *nabla* or del operator, that in Cartesian coordinates is given by

$$\boldsymbol{\nabla} = \mathbf{x}\frac{\partial}{\partial x} + \mathbf{y}\frac{\partial}{\partial y} + \mathbf{z}\frac{\partial}{\partial z} \quad \text{(B.13)}$$

B.2.1 Gradient

Application of the del operator to a scalar field, f, is known as the gradient, and is given in Cartesian, cylindrical, and spherical coordinates (see Figure B.1) as

$$\boldsymbol{\nabla} f = \frac{\partial f}{\partial x}\mathbf{x} + \frac{\partial f}{\partial y}\mathbf{y} + \frac{\partial f}{\partial z}\mathbf{z} \tag{B.14a}$$

$$= \frac{\partial f}{\partial \rho}\boldsymbol{\rho} + \frac{1}{\rho}\frac{\partial f}{\partial \phi}\boldsymbol{\phi} + \frac{\partial f}{\partial z}\mathbf{z} \tag{B.14b}$$

$$= \frac{\partial f}{\partial r}\mathbf{r} + \frac{1}{r}\frac{\partial f}{\partial \theta}\boldsymbol{\theta} + \frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\boldsymbol{\phi} \tag{B.14c}$$

Sometimes written grad f, the gradient represents the direction and magnitude of the maximum slope of the scalar field, f.

B.2.2 Divergence

Given a vector field, $\mathbf{A}$, the divergence is given by

$$\boldsymbol{\nabla} \cdot \mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \tag{B.15a}$$

$$= \frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho A_\rho\right) + \frac{1}{\rho}\frac{\partial A_\phi}{\partial \phi} + \frac{\partial A_z}{\partial z} \tag{B.15b}$$

$$= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 A_r\right) + \frac{1}{r\sin\theta}\frac{\partial}{\partial \theta}\left(A_\theta \sin\theta\right) + \frac{1}{r\sin\theta}\frac{\partial A_\phi}{\partial \phi} \tag{B.15c}$$

Also written div $\mathbf{A}$, the divergence represents the emergent (or outward-directed) flux density of a vector field. It can be defined as the limiting surface integral around an infinitesimally small volume of space,

$$\boldsymbol{\nabla} \cdot \mathbf{A} = \lim_{\Delta V \to 0} \frac{1}{\Delta V} \oiint_S \mathbf{A} \cdot d\mathbf{S} \tag{B.16}$$

B.2.3 Curl

The curl of a vector field, $\mathbf{A}$, is given in Cartesian coordinates by

$$\boldsymbol{\nabla} \times \mathbf{A} = \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right)\mathbf{x} + \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right)\mathbf{y} + \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right)\mathbf{z} \tag{B.17}$$

or in tensor index notation,

$$(\boldsymbol{\nabla} \times \mathbf{A})^i = \epsilon^{ijk} \frac{\partial A_k}{\partial x^j} \tag{B.18}$$

This can also be written in terms of the skew-symmetric matrix expansion of $\mathbf{A}$,

$$\boldsymbol{\nabla} \times \mathbf{A} = -[\dot{\mathbf{A}}]_\times \dot{\boldsymbol{\nabla}} \tag{B.19}$$

where the overdots indicate that $\boldsymbol{\nabla}$ differentiates $\mathbf{A}$ despite the postfix notation. In cylindrical coordinates, the curl is

$$\begin{aligned}\boldsymbol{\nabla} \times \mathbf{A} = \left(\frac{1}{\rho}\frac{\partial A_z}{\partial \phi} - \frac{\partial A_\phi}{\partial z}\right) \boldsymbol{\rho} + \left(\frac{\partial A_\rho}{\partial z} - \frac{\partial A_z}{\partial \rho}\right) \boldsymbol{\phi} \\ + \frac{1}{\rho}\left(\frac{\partial}{\partial \rho}(\rho A_\phi) - \frac{\partial A_\rho}{\partial \phi}\right) \mathbf{z}\end{aligned} \tag{B.20}$$

and in spherical coordinates it is given by

$$\begin{aligned}\boldsymbol{\nabla} \times \mathbf{A} = \frac{1}{r \sin\theta}\left[\frac{\partial}{\partial \theta}(A_\phi \sin\theta) - \frac{\partial A_\theta}{\partial \phi}\right] \mathbf{r} \\ + \frac{1}{r}\left[\frac{1}{\sin\theta}\frac{\partial A_r}{\partial \phi} - \frac{\partial}{\partial r}(r A_\phi)\right] \boldsymbol{\theta} + \frac{1}{r}\left[\frac{\partial}{\partial r}(r A_\theta) - \frac{\partial}{\partial \theta} A_r\right] \boldsymbol{\phi}\end{aligned} \tag{B.21}$$

Sometimes called the *rotational* and written as either curl $\mathbf{A}$ or rot $\mathbf{A}$, the curl represents the circulation density of a vector field, with direction given by the right-hand rule. The curl projected onto a plane perpendicular to an arbitrary unit vector $\mathbf{n}$ can be defined as the limiting closed path integral of the vector field in that plane enclosing an infinitesimally small area,

$$(\boldsymbol{\nabla} \times \mathbf{A}) \cdot \mathbf{n} = \lim_{\Delta A \to 0} \frac{1}{\Delta A} \oint_l \mathbf{A} \cdot d\mathbf{l} \tag{B.22}$$

B.2.4 Laplacian

The divergence of the gradient, in any coordinate system, is known as the *Laplacian* or *Laplace operator* of the given scalar field,

$$\boldsymbol{\nabla} \cdot \boldsymbol{\nabla} f = \boldsymbol{\nabla}^2 f \tag{B.23}$$

The Laplacian is proportional to the rate at which the average of a function over a spherical region surrounding a given point deviates from its value at that point as a function of the radius of that sphere.

The Laplacian of a scalar field in Cartesian, cylindrical, and spherical coordinates is

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \tag{B.24a}$$

$$= \frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho \frac{\partial f}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 f}{\partial \phi^2} + \frac{\partial^2 f}{\partial z^2} \tag{B.24b}$$

$$= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial f}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial f}{\partial \theta}\right) + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2 f}{\partial \phi^2} \tag{B.24c}$$

One may write the Laplacian of a vector field, $\mathbf{A}$, as well. In Cartesian coordinates, it is simply given by the Laplacians of the individual components,

$$\nabla^2 \mathbf{A} = \left(\nabla^2 A_x\right)\mathbf{x} + \left(\nabla^2 A_y\right)\mathbf{y} + \left(\nabla^2 A_z\right)\mathbf{z} \tag{B.25}$$

The form is not so simple in cylindrical or spherical coordinates, because the directions of the unit vectors themselves are also functions of position. The relevant expressions may be derived by use of the identity,

$$\nabla^2 \mathbf{A} = \nabla\left(\nabla \cdot \mathbf{A}\right) - \nabla \times \nabla \times \mathbf{A} \tag{B.26a}$$

$$= \left(\nabla^2 A_\rho - \frac{2}{\rho^2}\frac{\partial A_\phi}{\partial \phi} - \frac{A_\rho}{\rho^2}\right)\boldsymbol{\rho} + \left(\nabla^2 A_\phi + \frac{2}{\rho^2}\frac{\partial A_\rho}{\partial \phi} - \frac{A_\phi}{\rho^2}\right)\boldsymbol{\phi} + \left(\nabla^2 A_z\right)\mathbf{z} \tag{B.26b}$$

$$\begin{aligned} = &\left[\nabla^2 A_r - \frac{2}{r^2}\left(A_r + \cot\theta A_\theta + \csc\theta \frac{\partial A_\phi}{\partial \phi} + \frac{\partial A_\theta}{\partial \theta}\right)\right]\mathbf{r} \\ &+ \left[\nabla^2 A_\theta - \frac{1}{r^2}\left(\csc^2\theta A_\theta - 2\frac{A_r}{\partial \theta} + 2\frac{\cot\theta}{\sin\theta}\frac{\partial A_\phi}{\partial \phi}\right)\right]\boldsymbol{\theta} \\ &+ \left[\nabla^2 A_\phi - \frac{1}{r^2}\left(\csc^2\theta A_\phi - 2\csc\theta \frac{\partial A_r}{\partial \phi} - 2\frac{\cot\theta}{\sin\theta}\frac{\partial A_\theta}{\partial \phi}\right)\right]\boldsymbol{\phi} \end{aligned} \tag{B.26c}$$

B.3 IDENTITIES

For any vectors $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, and scalars f and g [2, 3],

$$(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{C}\times\mathbf{D}) = \mathbf{A}\cdot\mathbf{B}\times(\mathbf{C}\times\mathbf{D}) \tag{B.27a}$$

$$= \mathbf{A}\cdot(\mathbf{B}\cdot\mathbf{D}\mathbf{C} - \mathbf{B}\cdot\mathbf{C}\mathbf{D}) \tag{B.27b}$$

$$= (\mathbf{A}\cdot\mathbf{C})\,(\mathbf{B}\cdot\mathbf{D}) - (\mathbf{A}\cdot\mathbf{D})\,(\mathbf{B}\cdot\mathbf{C}) \tag{B.27c}$$

$$(\mathbf{A}\times\mathbf{B})\times(\mathbf{C}\times\mathbf{D}) = (\mathbf{A}\times\mathbf{B}\cdot\mathbf{D})\,\mathbf{C} - (\mathbf{A}\times\mathbf{B}\cdot\mathbf{C})\,\mathbf{D} \tag{B.28}$$

$$\boldsymbol{\nabla}(fg) = f\boldsymbol{\nabla}g + g\boldsymbol{\nabla}f \tag{B.29}$$

$$\boldsymbol{\nabla}\cdot(f\mathbf{A}) = f\boldsymbol{\nabla}\cdot\mathbf{A} + \mathbf{A}\cdot\boldsymbol{\nabla}f \tag{B.30}$$

$$\boldsymbol{\nabla}\times(f\mathbf{A}) = f\boldsymbol{\nabla}\times\mathbf{A} + \boldsymbol{\nabla}f\times\mathbf{A} \tag{B.31}$$

$$\boldsymbol{\nabla}(\mathbf{A}\cdot\mathbf{B}) = (\mathbf{A}\cdot\boldsymbol{\nabla})\mathbf{B} + (\mathbf{B}\cdot\boldsymbol{\nabla})\mathbf{A} + \mathbf{A}\times(\boldsymbol{\nabla}\times\mathbf{B}) + \mathbf{B}\times(\boldsymbol{\nabla}\times\mathbf{A}) \tag{B.32}$$

$$\boldsymbol{\nabla}\cdot(\mathbf{A}\times\mathbf{B}) = \mathbf{B}\cdot(\boldsymbol{\nabla}\times\mathbf{A}) - \mathbf{A}\cdot(\boldsymbol{\nabla}\times\mathbf{B}) \tag{B.33}$$

$$\boldsymbol{\nabla}\times(\mathbf{A}\times\mathbf{B}) = \mathbf{A}\,(\boldsymbol{\nabla}\cdot\mathbf{B}) - \mathbf{B}\,(\boldsymbol{\nabla}\cdot\mathbf{A}) + (\mathbf{B}\cdot\boldsymbol{\nabla})\,\mathbf{A} - (\mathbf{A}\cdot\boldsymbol{\nabla})\,\mathbf{B} \tag{B.34}$$

$$\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\mathbf{A} = \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{A}) - \boldsymbol{\nabla}^2\mathbf{A} \tag{B.35}$$

$$\boldsymbol{\nabla}\times\boldsymbol{\nabla}f = 0 \tag{B.36}$$

$$\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times\mathbf{A}) = 0 \tag{B.37}$$

B.4 INTEGRAL THEOREMS

Two important theorems in three-dimensional vector calculus are the *divergence theorem* and *Stokes theorem*, illustrated in Figure B.2. The divergence theorem states that the integral of the divergence of a vector field over a finite volume is given by the closed surface integral of the flux crossing the boundary of that volume,

$$\iiint_V (\boldsymbol{\nabla}\cdot\mathbf{F})dV = \oint\!\!\!\!\oint_S \mathbf{F}\cdot d\mathbf{S} \tag{B.38}$$

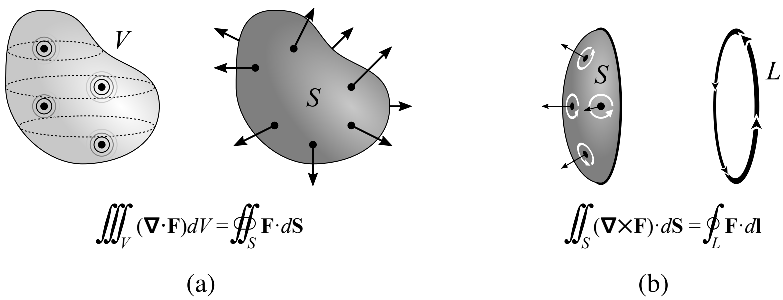


Figure B.2 Graphical interpretation of (a) the divergence theorem, and (b) Stokes theorem.

Stokes theorem states that the integral of the curl of a vector field on a finite surface is equal to the closed path integral around the loop bounding that surface,

$$\iint_S (\nabla \times \mathbf{F}) \cdot d\mathbf{S} = \oint_L \mathbf{F} \cdot d\mathbf{l} \tag{B.39}$$

In these equations, $d\mathbf{S}$ is interpreted as $\mathbf{n}dS$ where $\mathbf{n}$ is the outward-directed unit normal of the surface S, and $d\mathbf{l} = \mathbf{u}dl$ is the tangential unit vector on the curve L. Both of these theorems (and many more) are special cases of a much broader fundamental theorem of calculus in spacetime algebra, which was described in Section 8.1.2.

References

[1] G. Arfken and H. Weber, *Mathematical Methods for Physicists*, 4th ed. San Diego, CA: Academic Press, 1995.

[2] C. Balanis, *Advanced Engineering Electromagnetics*. New York: Wiley, 1989.

[3] M. A. Morgan, *Principles of RF and Microwave Design*. Norwood, MA: Artech House, November 2019.

Appendix C

Ricci Calculus

Ricci calculus [1–3] is a system for performing multidimensional calculations using mathematical objects called tensors. Tensors may loosely be considered arrays of numbers like vectors or matrices and, as such, tensor equations intrinsically represent coupled systems of equations in real numbers.

C.1 INDICES

A tensor variable is generally indicated by the presence of indices in either the superscript or subscript position, or both. An index in the superscript position is not intended to indicate exponentiation, rather it signifies the relationship elements within the array have to one another according to the metric signature used.

C.1.1 Index Range

Indices written with lowercase Greek letters in either position are assumed to range from 0 to 3, where, by convention, index 0 refers to the temporal component and indices $1 - 3$ denote the three spatial coordinates. Thus, a spacetime four-position may be written as

$$X^{\mu} = \left(X^0, X^1, X^2, X^3\right) = (ct, x, y, z) \tag{C.1}$$

Indices written with lowercase roman letters range from 1 to 3 only, thereby isolating spatial components,

$$X^{k} = \left(X^1, X^2, X^3\right) = (x, y, z) \tag{C.2}$$

C.1.2 Rank

Tensors having only one index are called first-rank tensors. Those having two indices are second-rank, and so on. There is no limit to the rank that a tensor may have. First-rank tensors may generally be thought of as four-vectors in spacetime, such as the four-position, four-velocity, four-wavevector, and four-current density, below,

$$X^\mu = \left(X^0, X^1, X^2, X^3\right) = (ct, x, y, z) \tag{C.3a}$$

$$U^\mu = \left(U^0, U^1, U^2, U^3\right) = (c, v_x, v_y, v_z)\cosh\zeta \tag{C.3b}$$

$$K^\mu = \left(K^0, K^1, K^2, K^3\right) = \left(\tfrac{\omega}{c}, k_x, k_y, k_z\right) \tag{C.3c}$$

$$J^\mu = \left(J^0, J^1, J^2, J^3\right) = (c\rho, J_x, J_y, J_z) \tag{C.3d}$$

Second-rank tensors more closely resemble matrices, and include such quantities as the Faraday field tensor.

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix} \tag{C.4}$$

The order of indices is significant. For example, the Faraday tensor above is antisymmetric, so that $F^{\mu\nu} = -F^{\nu\mu}$.

C.1.3 Contravariant and Covariant Indices

Indices in the superscript position are referred to as contravariant, while those in the subscript position are called covariant. Any tensor, of any rank, can be written with any combination of contravariant and covariant indices. When using the (+ − − −) metric signature, raising or lowering indices results in the negation of all elements associated with the spatial dimensions. Thus,

$$X_\mu = (X_0, X_1, X_2, X_3) = (ct, -x, -y, -z) \tag{C.5}$$

and

$$F^\mu{}_\nu = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ E_x/c & 0 & B_z & -B_y \\ E_y/c & -B_z & 0 & B_x \\ E_z/c & B_y & -B_x & 0 \end{pmatrix} \tag{C.6}$$

C.2 TENSOR ADDITION AND MULTIPLICATION

Tensors of the same rank may be added much like vectors and matrices by summing the elements from both having the same index values. Thus,

$$X^{\mu} + Y^{\mu} = \left(X^0 + Y^0, X^1 + Y^1, X^2 + Y^2, X^3 + Y^3\right) \tag{C.7}$$

Tensors of any rank can be multiplied, the result of which is dependent on the position and pairing of indices, as described below.

C.2.1 Outer Product

When two tensors are multiplied that have no indices in common, the product inherits the indices from its constituents and has rank equal to the sum of their ranks.

$$X^{\mu}Y^{\nu} = Z^{\mu\nu} = \begin{pmatrix} X^0Y^0 & X^0Y^1 & X^0Y^2 & X^0Y^3 \\ X^1Y^0 & X^1Y^1 & X^1Y^2 & X^1Y^3 \\ X^2Y^0 & X^2Y^1 & X^2Y^2 & X^2Y^3 \\ X^3Y^0 & X^3Y^1 & X^3Y^2 & X^3Y^3 \end{pmatrix} \tag{C.8}$$

The element tensors may be of any rank, with any combination of contravariant and covariant indices. When both tensors are first-rank, this is effectively the dyadic product from vector calculus introduced in Section B.1.4.

The product of a tensor with a scalar simply results in the scalar multiplying each element of the tensor. This is consistent with the above outer product if the scalar is treated as a zero-rank tensor.

C.2.2 Inner Product

When two tensors are multiplied having the same index label, and one is contravariant while the other is covariant, the Einstein summation rule is activated, triggering an implicit summation over the repeated index,

$$Y^{\mu} = L^{\mu}{}_{\nu}X^{\nu} = \sum_{\nu=0}^{3} L^{\mu}{}_{\nu}X^{\nu} = \begin{pmatrix} L^0{}_0X^0 + L^0{}_1X^1 + L^0{}_2X^2 + L^0{}_3X^3 \\ L^1{}_0X^0 + L^1{}_1X^1 + L^1{}_2X^2 + L^1{}_3X^3 \\ L^2{}_0X^0 + L^2{}_1X^1 + L^2{}_2X^2 + L^2{}_3X^3 \\ L^3{}_0X^0 + L^3{}_1X^1 + L^3{}_2X^2 + L^3{}_3X^3 \end{pmatrix} \tag{C.9}$$

Note that the summed index, ν, is absent from the tensor result, Y^{μ}. The index is said to have been contracted. More than one index can be contracted at a time, each triggering its own summation. The product has rank equal to the number of remaining, uncontracted indices.

The square norm of a four-vector is found by contracting the indices of its contravariant and covariant forms. For example,

$$U_\mu U^\mu = c^2 \tag{C.10}$$

Expressions wherein an index is repeated, but in the same position (both contravariant or both covariant, whether or not on separate tensors), are considered ill-formed. Similarly, matched indices on both sides of an equation must appear in the same position. For example,

$$Y^\mu = L^\mu{}_\nu X^\nu \tag{C.11}$$

is a well-formed tensor equation, while

$$Y_\mu \stackrel{?}{=} L^\mu{}_\nu X^\nu \tag{C.12}$$

is not acceptable.

C.2.3 Index Manipulation

Indices may be raised or lowered by multiplication with the metric, a second-rank tensor given in contravariant form by

$$\eta^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{C.13}$$

The covariant form is identical,

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{C.14}$$

The signs of the diagonal elements of the metric tensor are given by the metric signature, in this case (+ – – –). Two examples of using the metric tensor to raise or lower indices are as follows

$$X^\mu = \eta^{\mu\nu} X_\nu \tag{C.15a}$$

$$F_\mu{}^\nu = \eta_{\mu\sigma} F^{\sigma\nu} \tag{C.15b}$$

C.3 DIFFERENTIATION

The principal differential operator in tensor calculus, the spacetime gradient, behaves like a first-rank tensor,

$$\partial_\mu = \left(\frac{\partial}{\partial X^0}, \frac{\partial}{\partial X^1}, \frac{\partial}{\partial X^2}, \frac{\partial}{\partial X^3}\right) = \left(\frac{1}{c}\frac{\partial}{\partial t}, \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) \tag{C.16}$$

A compact way to write this is

$$\partial_\mu = (\partial_0, \boldsymbol{\nabla}) \tag{C.17}$$

Contracting the indices of this operator with another vector is called the spacetime divergence. For example,

$$\partial_\mu J^\mu = \frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{J} \tag{C.18}$$

As an operator, the spacetime gradient has no square norm, per se, but its inner product with itself does form a second-order, scalar differential operator known as the d'Alembertian [4],

$$\partial_\mu \partial^\mu = \partial_0^2 - \boldsymbol{\nabla}^2 = \Box^2 \tag{C.19}$$

C.4 LEVI-CIVITA SYMBOL

A special object that is useful in certain tensor equations is the Levi-Civita symbol. Not strictly a tensor, because it does not transform between reference frames as tensors do, it nevertheless has indices and interacts with other tensors in the same manner. Written as a lowercase epsilon (ϵ), it can have any rank, and is completely antisymmetric in all dimensions. Its elements are defined as -1 when the indices represent an even permutation of consecutive numerical order (0123), $+1$ if they represent an odd permutation, and 0 otherwise,

$$\epsilon^{ijkl} = \begin{cases} -1 & \text{if } (i,j,k,l) \text{ is } (0,1,2,3) \text{ or any even permutation thereof} \\ +1 & \text{if } (i,j,k,l) \text{ is an odd permutation of } (0,1,2,3) \\ 0 & \text{otherwise} \end{cases} \tag{C.20}$$

An even permutation is one which can be obtained by an even number of transpositions of elements, whether or not those elements are adjacent. Odd permutations are those for which the required number of transpositions is odd. Any repeated indices automatically make the value 0.

The Levi-Civita symbols of various ranks can be used in combination with the Einstein summation rule to write vector cross products, triple products, and curls in compact form, as well as to form the Hodge dual [5] of certain tensor quantities,

$$G^{\mu\nu} = \star F^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\sigma\tau}F_{\sigma\tau} \tag{C.21}$$

C.5 SPECIAL NOTATIONS

Tensor index notation is equipped with a few standard shorthand forms to make equations more compact [6].

C.5.1 Symmetrization

Indices enclosed in parentheses, even across multiple tensors in a product, represent the *symmetrization* of the tensor or tensor product with respect to those indices. In other words, the tensor or product is replicated for all possible permutations of the indices, and the results averaged,

$$T_{(\mu\nu)} = \frac{1}{2}\left(T_{\mu\nu} + T_{\nu\mu}\right) \tag{C.22a}$$

$$a^{(\mu}b^{\nu)} = \frac{1}{2}\left(a^{\mu}b^{\nu} + a^{\nu}b^{\mu}\right) \tag{C.22b}$$

C.5.2 Antisymmetrization

Indices enclosed in square brackets, even across multiple tensors in a product, represent the antisymmetrization of the tensor or tensor product with respect to those indices. In this case, the tensor or product is replicated and averaged over all signed permutations, wherein the sign is given by the parity of the permutation (negative for odd permutations). For example,

$$\partial^{[\mu} A^{\nu]} = \frac{1}{2} \left(\partial^\mu A^\nu - \partial^\nu A^\mu \right) \tag{C.23a}$$

$$\partial_{[\mu} F_{\nu\sigma]} = \frac{1}{6} \left(\partial_\mu F_{\nu\sigma} + \partial_\sigma F_{\mu\nu} + \partial_\nu F_{\sigma\mu} - \partial_\mu F_{\sigma\nu} - \partial_\sigma F_{\nu\mu} - \partial_\nu F_{\mu\sigma} \right) \tag{C.23b}$$

C.5.3 Differentiated Index

A comma preceding an index denotes differentiation with respect to that index,

$$\phi_{,\mu} = \partial_\mu \phi \tag{C.24a}$$

$$G^{\mu\nu}{}_{,\nu} = \partial_\nu G^{\mu\nu} \tag{C.24b}$$

References

[1] Wikipedia. (2023, December) Ricci calculus. https://en.wikipedia.org/wiki/Ricci_calculus.

[2] I. L. Shapiro, *A Primer in Tensor Analysis and Relativity*. Switzerland: Springer, 2019.

[3] J. M. Charap, *Covariant Electrodynamics: A Concise Guide*. Maryland: Johns Hopkins University Press, 2011.

[4] M. A. Morgan, *Principles of RF and Microwave Design*. Norwood, MA: Artech House, November 2019.

[5] Wikipedia. (2024, January) Hodge star operator. https://en.wikipedia.org/wiki/Hodge_star_operator.

[6] L. Susskind and A. Friedman, *Special Relativity and Classical Field Theory: The Theoretical Minimum*. New York: Basic Books, 2017.

Appendix D

Spacetime Algebra

Spacetime algebra [1–10] is a geometric, or Clifford algebra, based on the four-dimensional Minkowski space spanned by the basis vectors γ_μ for $\mu = 0, 1, 2, 3$, having a square norm for metric signature (+ − − −) as follows,

$$\gamma_\mu^2 = \begin{cases} +1 & \mu = 0 \\ -1 & \mu = 1, 2, 3 \end{cases} \tag{D.1}$$

D.1 ALGEBRAIC DOMAIN

The complete domain of objects within this algebra may be written in terms of the multivector basis,

$$\{1\}, \{\gamma_\mu\}, \{\sigma_k, i\sigma_k\}, \{i\gamma_\mu\}, \{i\} \tag{D.2}$$

where

$$\mu = 0, 1, 2, 3 \tag{D.3a}$$

$$k = 1, 2, 3 \tag{D.3b}$$

In this set, 1 is the scalar basis element, γ_μ are the vectors, σ_k and $i\sigma_k$ are the bivectors, $i\gamma_\mu$ are the pseudovectors, and i is the pseudoscalar ($i^2 = -1$). Some useful expansions of these elements follow,

$$\sigma_k = \gamma_k\gamma_0 \tag{D.4a}$$

$$i = \gamma_0\gamma_1\gamma_2\gamma_3 = \sigma_1\sigma_2\sigma_3 \tag{D.4b}$$

Table D.1

Commutation Signatures for All Basis Element Pairs

	1	γ_0	γ_1	γ_2	γ_3	σ_1	σ_2	σ_3	$i\sigma_1$	$i\sigma_2$	$i\sigma_3$	$i\gamma_0$	$i\gamma_1$	$i\gamma_2$	$i\gamma_3$	i
1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
γ_0	+	+	−	−	−	−	−	−	+	+	+	−	+	+	+	−
γ_1	+	−	+	−	−	−	+	+	+	−	−	+	−	+	+	−
γ_2	+	−	−	+	−	+	−	+	−	+	−	+	+	−	+	−
γ_3	+	−	−	−	+	+	+	−	−	−	+	+	+	+	−	−
σ_1	+	−	−	+	+	+	−	−	+	−	−	−	−	+	+	+
σ_2	+	−	+	−	+	−	+	−	−	+	−	−	+	−	+	+
σ_3	+	−	+	+	−	−	−	+	−	−	+	−	+	+	−	+
$i\sigma_1$	+	+	+	−	−	+	−	−	+	−	−	+	+	−	−	+
$i\sigma_2$	+	+	−	+	−	−	+	−	−	+	−	+	−	+	−	+
$i\sigma_3$	+	+	−	−	+	−	−	+	−	−	+	+	−	−	+	+
$i\gamma_0$	+	−	+	+	+	−	−	−	+	+	+	+	−	−	−	−
$i\gamma_1$	+	+	−	+	+	−	+	+	+	−	−	−	+	−	−	−
$i\gamma_2$	+	+	+	−	+	+	−	+	−	+	−	−	−	+	−	−
$i\gamma_3$	+	+	+	+	−	+	+	−	−	−	+	−	−	−	+	−
i	+	−	−	−	−	+	+	+	+	+	+	−	−	−	−	+

We can also write the following equivalents,

$$i\sigma_1 = \sigma_1 i = \gamma_3\gamma_2 = \sigma_2\sigma_3 \tag{D.5a}$$

$$i\sigma_2 = \sigma_2 i = \gamma_1\gamma_3 = \sigma_3\sigma_1 \tag{D.5b}$$

$$i\sigma_3 = \sigma_3 i = \gamma_2\gamma_1 = \sigma_1\sigma_2 \tag{D.5c}$$

$$i\gamma_0 = -\gamma_0 i = \gamma_3\gamma_2\gamma_1 \tag{D.5d}$$

$$i\gamma_1 = -\gamma_1 i = \gamma_0\gamma_3\gamma_2 \tag{D.5e}$$

$$i\gamma_2 = -\gamma_2 i = \gamma_3\gamma_0\gamma_1 \tag{D.5f}$$

$$i\gamma_3 = -\gamma_3 i = \gamma_2\gamma_1\gamma_0 \tag{D.5g}$$

Scalars are called grade-0 objects, while vectors are grade-1, bivectors are grade-2, and so on. For convenience, the commutation signatures[1] of all pairs of basis elements are given in Table D.1. In the sections that follow, unless otherwise noted,

1 That is, for any two elements a and b, the commutation signature is positive if $ab = ba$, and negative if $ab = -ba$.

- ϕ, ψ, $\chi \ldots$ are scalars;
- γ_μ, σ_μ, ρ_μ, $\phi_\mu \ldots$ are basis vectors;
- a, b, $c \ldots$ are spacetime vectors (grade-1);
- K, L, $M \ldots$ are homogeneous multivectors (having grades k, l, m, etc.);
- F and G in particular will represent bivectors (i.e., grade $f = g = 2$);
- $\mathcal{M}$, $\mathcal{N}$, $\mathcal{O} \ldots$ are mixed-grade multivectors.

D.2 GRADE PROJECTION

The notation $\langle \mathcal{M} \rangle_p$ returns the grade-p component of the multivector $\mathcal{M}$, called the projection. Thus, a complete multivector may be written as the sum of its projected components,

$$\mathcal{M} = \sum_{p=0}^{n} \langle \mathcal{M} \rangle_p \tag{D.6}$$

Grade-0 projection is sometimes called the *trace*, and is often written without the subscript,

$$\langle \mathcal{M} \rangle_0 = \langle \mathcal{M} \rangle \tag{D.7}$$

As a special case, $\langle \mathcal{M} \rangle_i$ shall represent the scalar and pseudoscalar (grade-0 and grade-4) components of $\mathcal{M}$, which are invariant under Lorentz transformation. I will sometimes use the notation $\mathcal{M}_{k,l,m\ldots}$ for multivectors having only k, l, and m-grade components, and furthermore use $\mathcal{M}_{even}$ and $\mathcal{M}_{odd}$ to refer to multivectors having only even and odd-grade components, respectively.

D.3 PRODUCTS

D.3.1 Geometric Product

The geometric product of two homogeneous components, K and L, is written simply KL, and has the following expansion

$$KL = \langle KL \rangle_{k+l} + \langle KL \rangle_{k+l-2} + \cdots + \langle KL \rangle_{|k-l|} \tag{D.8}$$

where any projection to a grade $(k + l) > 4$ is zero by definition (such objects would compose more than four linearly independent vectors, an impossibility in four-dimensional space).

D.3.2 Dot and Wedge Products

The dot and wedge product notations are reserved for the lowest and highest-grade components of the summation (D.8), such that

$$K \cdot L = \langle KL \rangle_{|k-l|} \tag{D.9}$$

$$K \wedge L = \langle KL \rangle_{k+l} \tag{D.10}$$

with the exception that, when one object is a scalar, there is only one term. By convention, we consider any product with a scalar (grade-0) as a wedge product rather than a dot product,

$$\psi \wedge K = \langle \psi K \rangle_k = \psi K \tag{D.11a}$$

$$\psi \cdot K = 0 \tag{D.11b}$$

This convention avoids double counting products of scalars as both dot and wedge products in certain identities, while preserving the expected symmetries of products with even-grade components.

Thus, for the special case in which one element is a grade-1 vector, we have

$$aK = a \cdot K + a \wedge K \tag{D.12}$$

$$Ka = K \cdot a + K \wedge a \tag{D.13}$$

where

$$a \cdot K = (-1)^{k+1} K \cdot a = \frac{1}{2}\left(aK + (-1)^{k+1} Ka\right) \tag{D.14a}$$

$$a \wedge K = (-1)^{k} K \wedge a = \frac{1}{2}\left(aK + (-1)^{k} Ka\right) \tag{D.14b}$$

By extension, then, for a general multivector, $\mathcal{M}$, we have

$$a\mathcal{M} = a \cdot \mathcal{M} + a \wedge \mathcal{M} \tag{D.15a}$$

$$\mathcal{M}a = \mathcal{M} \cdot a + \mathcal{M} \wedge a \tag{D.15b}$$

When both elements are vectors,

$$ab = a \cdot b + a \wedge b \tag{D.16a}$$

$$a \cdot b = b \cdot a = \frac{1}{2}(ab + ba) \tag{D.16b}$$

$$a \wedge b = -b \wedge a = \frac{1}{2}(ab - ba) \tag{D.16c}$$

The wedge product is associative, while the dot product in general is not.

D.3.3 Commutation Products

The commutator bracket is defined for any $\mathcal{M}$ and $\mathcal{N}$ as follows,

$$[\mathcal{M}, \mathcal{N}] = -[\mathcal{N}, \mathcal{M}] = \tfrac{1}{2}(\mathcal{M}\mathcal{N} - \mathcal{N}\mathcal{M}) \tag{D.17}$$

For grade-1 vectors times other single-grade objects, the commutator bracket is equivalent to either the dot or wedge product, depending on the grade,

$$[a, K] = \begin{cases} a \cdot K & k \text{ is even} \\ a \wedge K & k \text{ is odd} \end{cases} \tag{D.18}$$

The commutator bracket with a bivector F is grade-preserving,

$$\langle [F, \mathcal{M}] \rangle_k = [F, \langle \mathcal{M} \rangle_k] \tag{D.19}$$

The anticommutator bracket is defined as

$$\{\mathcal{M}, \mathcal{N}\} = \{\mathcal{N}, \mathcal{M}\} = \tfrac{1}{2}(\mathcal{M}\mathcal{N} + \mathcal{N}\mathcal{M}) \tag{D.20}$$

and has the following properties

$$\{a, K\} = \begin{cases} a \wedge K & k \text{ is even} \\ a \cdot K & k \text{ is odd} \end{cases} \tag{D.21}$$

For bivectors, F and G, the commutator and anticommutator products are closely related to the traditional cross and dot products, respectively, from classical vector calculus,

$$[F, G]\, i^{-1} = F \times G \tag{D.22a}$$

$$\{F, G\} = F \cdot G \tag{D.22b}$$

D.4 GEOMETRIC PROJECTION

The projection of a multivector $\mathcal{M}$ onto a k-vector K is written

$$\mathcal{M}_K = (\mathcal{M} \cdot K)K^{-1} \tag{D.23}$$

and the rejection of $\mathcal{M}$ from K is

$$\mathcal{M}_{\not K} = (\mathcal{M} \wedge K)K^{-1} \tag{D.24}$$

It is useful in some special cases to have an explicit notation for the following reversed forms, called *preprojection* and *prerejection*, respectively,

$$_K\mathcal{M} = K^{-1}(K \cdot \mathcal{M}) \tag{D.25a}$$

$$_{\not K}\mathcal{M} = K^{-1}(K \wedge \mathcal{M}) \tag{D.25b}$$

These are equivalent to $\mathcal{M}_K$ and $\mathcal{M}_{\not K}$ above unless K is a nonsimple bivector, an irreducible complexity of grade-2 objects that shall be discussed in Section D.5.4.

Finally, note that all multivectors are the sum of their projections and rejections from a given vector, and that all vectors are the sum of their projections and rejections from a given homogeneous multivector,

$$\mathcal{M} = \mathcal{M}_a + \mathcal{M}_{\not a} \tag{D.26a}$$

$$a = a_K + a_{\not K} = {}_K a + {}_{\not K} a \tag{D.26b}$$

D.5 STRUCTURAL CHARACTERISTICS

Certain unary operations that alter the structure of a general multivector are made clearer if we write that multivector as the sum of its graded elements,

$$\mathcal{M} = \alpha + v + F + iw + i\beta \tag{D.27}$$

where

$$\alpha = \langle \mathcal{M} \rangle_0 \tag{D.28a}$$

$$v = \langle \mathcal{M} \rangle_1 \tag{D.28b}$$

$$F = \langle \mathcal{M} \rangle_2 \tag{D.28c}$$

$$iw = \langle \mathcal{M} \rangle_3 \tag{D.28d}$$

$$i\beta = \langle \mathcal{M} \rangle_4 \tag{D.28e}$$

In these expressions, α and β are scalars, v and w are four-vectors, F is a bivector, and i, as usual, is the unit pseudoscalar. It is sometimes useful to think of $\alpha + i\beta$ as a complex scalar and $v + iw$ as a complex vector.

D.5.1 Duality

Right multiplication with i^{-1} produces the Hodge dual of a given multivector,

$$\star\mathcal{M} = \mathcal{M}i^{-1} = -i\alpha + iv - iF - w + \beta \tag{D.29}$$

Thus, the dual of a k-vector, K, is a component of grade $(4 - k)$ given up to a sign by iK or Ki. Note that i commutes with even-grade objects, while it anticommutes with odd-grade objects,

$$iK = (-1)^k Ki \tag{D.30}$$

Right multiplication of a complex vector with a timelike axis, γ_0, representing the worldline of a particular reference frame, produces the frame-dependent spacetime split of that vector.

$$(v + iw)\gamma_0 = \left(v^0 + iw^0\right) + (\mathbf{v} + i\mathbf{w}) \tag{D.31}$$

The spacetime split is a paravector containing both a scalar, representing the apparent temporal component of the four-vector in the given reference frame, and a bivector, representing the spatial three-vector component. These are both elements of the even subalgebra of spacetime. Right multiplication of elements of the even subalgebra with γ_0 restores the original complex four-vector forms.

$$\left(v^0 + iw^0\right)\gamma_0 + (\mathbf{v} + i\mathbf{w})\gamma_0 = v + iw \tag{D.32}$$

This can be referred to as reference-frame duality; paravectors and four-vectors are dual representations of the physical quantities in Minkowski spacetime [1].

D.5.2 Reversion

The reversion of a multivector $\mathcal{M}$, denoted $\widetilde{\mathcal{M}}$, is the reverse ordering of all vector products in its composition. Reversion thus has no effect on scalars or grade-1 vectors, and is similarly inactive upon pseudoscalars, but negates bivectors and trivectors; that is,

$$\widetilde{\mathcal{M}} = \alpha + v - F - iw + i\beta \tag{D.33}$$

Reversion operates recursively on all multivector products; for example,

$$(\mathcal{M}\mathcal{N})\tilde{} = \widetilde{\mathcal{N}}\widetilde{\mathcal{M}} \tag{D.34}$$

D.5.3 Magnitude

The reversed square of any k-vector can be written

$$K\widetilde{K} = \widetilde{K}K = \epsilon_K |K|^2 \tag{D.35}$$

where $|K|^2$ is the (scalar) positive square magnitude, and ϵ_K is its signature. The signature for each grade can be constrained as follows,

$$\epsilon_K = \begin{cases} 1 & k = 0 \\ \pm 1 & k = 1 \\ e^{i\psi} & k = 2 \\ \pm 1 & k = 3 \\ -1 & k = 4 \end{cases} \tag{D.36}$$

Note that the signature always has unit magnitude, and is real for all k-vectors except bivectors. The signature of a vector or pseudovector is either positive or negative depending on whether the associated vector is timelike or spacelike. For bivectors, the signature is complex.

The reversed square thus has (at most) a scalar and pseudoscalar part, and the positive square magnitude is merely the root-sum-squares of these two parts,

$$|K|^2 = \sqrt{\left|\left\langle K\widetilde{K}\right\rangle_0\right|^2 + \left|i\left\langle K\widetilde{K}\right\rangle_4\right|^2} \tag{D.37}$$

D.5.4 Simple Bivectors and the Polar Canonical Form

Simple bivectors are those that can be written as a single wedge product between vectors. The square norm of a bivector and its reversed square (K^2 and $\widetilde{K}K$, respectively) are complex in general, but scalar if the bivector is simple.

Any bivector can be written in a polar form,

$$F = \mathbf{f}e^{i\varphi} \tag{D.38}$$

where $\mathbf{f}$ is a simple bivector with norm $\mathbf{f}^2 = |F|^2$, and φ is a scalar, given by

$$\varphi = \frac{1}{2}\tan^{-1}\left(\frac{|F|^2\sin(2\varphi)}{|F|^2\cos(2\varphi)}\right) = \frac{1}{2}\tan^{-1}\left(\frac{\left\langle F^2\right\rangle_4}{i\left\langle F^2\right\rangle_0}\right) \tag{D.39a}$$

$$\therefore \mathbf{f} = Fe^{-i\varphi} \tag{D.39b}$$

When F is null ($F^2 = 0$), then $\mathbf{f}$ is also null. The phase term is then indeterminate; any value will do. Otherwise, the polar decomposition is unique and frame-independent [1, 3]. The same cannot be said of the Cartesian decomposition $F = \mathbf{a} + i\mathbf{b}$, which relies on a chosen worldline to separate the real and imaginary parts.

D.5.5 Conjugation

Calculations are facilitated by having a frame-independent conjugation which operates specifically on the polar form of the bivector, negating i wherever it appears,

$$\mathcal{M}^* = \alpha + v + \mathbf{f}e^{-i\varphi} - iw - i\beta \tag{D.40}$$

Note that while we allow $F = \frac{1}{c}\mathbf{E} + i\mathbf{B}$ in a particular reference frame, in general $F^* \neq \frac{1}{c}\mathbf{E} - i\mathbf{B}$.

D.5.6 Relative Reversion

In contrast, frame-dependent conjugation is provided by an operation called relative reversion, where the multivector is subjected to a spacetime split, reversed, and then unsplit, as follows,

$$\mathcal{M}^\dagger = \gamma_0\widetilde{\mathcal{M}}\gamma_0 = \alpha + \left(v^0 - \mathbf{v}\right)\gamma_0 + \left(\tfrac{1}{c}\mathbf{E} - i\mathbf{B}\right) + i\left(w^0 - \mathbf{w}\right)\gamma_0 - i\beta \tag{D.41}$$

Table D.2
Unary Operations on Multivectors

Name/Operation	Notation	Expression
Original form	$\mathcal{M}$	$\alpha + v + F + iw + i\beta$
Canonical form	$\mathcal{M}$	$\alpha + v + \mathbf{f}e^{i\varphi} + iw + i\beta$
Split form	$\mathcal{M}$	$\alpha + \left(v^0 + \mathbf{v}\right)\gamma_0 + \left(\frac{1}{c}\mathbf{E} + i\mathbf{B}\right) + i\left(w^0 + \mathbf{w}\right)\gamma_0 + i\beta$
Hodge dual	$\star\mathcal{M}$	$-i\alpha + iv - iF - w + \beta$
Reversion	$\widetilde{\mathcal{M}}$	$\alpha + v - F - iw + i\beta$
Conjugation	$\mathcal{M}^*$	$\alpha + v + \mathbf{f}e^{-i\varphi} - iw - i\beta$
Relative reversion	$\mathcal{M}^\dagger$	$\alpha + \left(v^0 - \mathbf{v}\right)\gamma_0 + \left(\frac{1}{c}\mathbf{E} - i\mathbf{B}\right) + i\left(w^0 - \mathbf{w}\right)\gamma_0 - i\beta$

This has the effect of reverse-ordering the relative vectors in the even subalgebra,

$$\left(\sigma_j \sigma_k\right)^\dagger = \sigma_k \sigma_j \tag{D.42}$$

Relative reversion can be used to write a magnitude-angle formula for the dot product akin to that known from Euclidean vector algebra,

$$a \cdot b = \sqrt{\left(a^\dagger \cdot a\right)\left(b^\dagger \cdot b\right)} \cos\theta_{a^\dagger b} \tag{D.43}$$

where $\theta_{a^\dagger b}$ is the angle between vectors $a^\dagger$ and b.

A summary of the unary operations applied to general multivectors is given in Table D.2. Reversion, conjugation, and relative reversion are all *involutions* of the algebra, meaning that each operation is its own inverse. Hodge duality is not; two applications result in the overall negation of its operand.

D.5.7 Multiplicative Inverse

The general multiplicative inverse of a k-vector can be written with the help of either the reversion or conjugation operators,

$$K^{-1} = \frac{\widetilde{K}}{\widetilde{K}K} = \frac{K^*}{K^*K} \tag{D.44}$$

Both forms are equivalent, so long as one is prepared to handle the complex denominator for bivectors ($\widetilde{F}F$) when using the reversion formula,

$$\frac{\widetilde{F}}{\widetilde{F}F} = \frac{-\mathbf{f}e^{i\varphi}}{(-\mathbf{f}e^{i\varphi})\,(\mathbf{f}e^{i\varphi})} = \frac{\mathbf{f}e^{i\varphi}}{\mathbf{f}^2 e^{i2\varphi}} = \frac{\mathbf{f}e^{-i\varphi}}{|F|^2} = \frac{F^*}{F^*F} \tag{D.45}$$

Since K^*K always returns a scalar for any k-vector, the inverse formula based on conjugation may be easier to apply in practice.

D.6 DIFFERENTIATION

The principal differentiation operator is the spacetime gradient,

$$\Box = \gamma^\mu \partial_\mu = \gamma^0\partial_0 + \gamma^1\partial_1 + \gamma^2\partial_2 + \gamma^3\partial_3 \tag{D.46}$$

D.6.1 Differential Primitives

Differentiation using the spacetime gradient is made simpler by the enumeration of certain primitive forms — forms from which more complex derivatives can then be derived. In this section only, $x = x^0\gamma_0 + x^1\gamma_1 + x^2\gamma_2 + x^3\gamma_3$ is the independent four-vector over which the derivative is taken, while a and K are constants, not dependent on x.

$$\Box x = \Box \cdot x = 4 \tag{D.47}$$

$$\Box \wedge x = 0 \tag{D.48}$$

$$\Box x^2 = \Box |x|^2 = 2x \tag{D.49}$$

$$\Box |x|^k = k|x|^{k-2}x \tag{D.50}$$

$$\Box \left(\frac{x}{|x|^k}\right) = \frac{4-k}{|x|^k} \tag{D.51}$$

$$\Box(x \cdot a) = (a \cdot \Box)x = a \tag{D.52}$$

$$\Box(x \cdot K) = (K \cdot \Box)x = kK \tag{D.53}$$

$$\Box(x \wedge K) = (K \wedge \Box)x = (4-k)K \tag{D.54}$$

$$\Box(Kx) = (-1)^k(4-2k)K \tag{D.55}$$

$$\Box e^{x\cdot a} = ae^{x\cdot a} \tag{D.56}$$

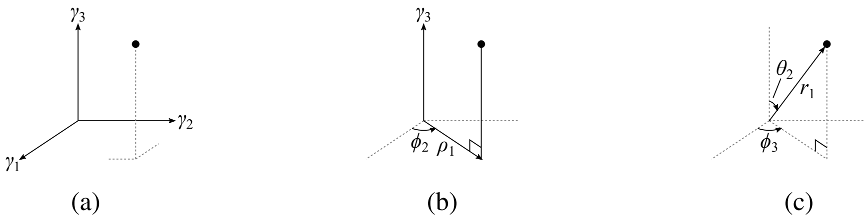


Figure D.1 Definition of spatial coordinate systems for spacetime algebra. (a) Cartesian. (b) Cylindrical. (c) Spherical. In all cases, the temporal axis, γ_0, is implicit.

$$\Box e^{x\cdot K} = kKe^{x\cdot K} \tag{D.57}$$

$$\Box \cdot \left(Ke^{x\cdot a}\right) = a \cdot \left(Ke^{x\cdot a}\right) \tag{D.58}$$

$$\Box \wedge \left(Ke^{x\cdot a}\right) = a \wedge \left(Ke^{x\cdot a}\right) \tag{D.59}$$

D.7 ALTERNATIVE COORDINATE SYSTEMS

Although general theoretical results can be derived in a coordinate-free manner, engineering problems tied to specific boundary conditions require the use of coordinate systems compatible with those boundaries. The basis vectors defined in Section D.1 are intrinsically Cartesian, as in Figure D.1(a), and are best suited to problems with rectangular boundaries. Techniques for incorporating other types of coordinates in the language of spacetime algebra are described here.

D.7.1 Cylindrical Coordinates

Our basis vectors for Cartesian coordinates were numbered γ_0 through γ_3, representing one temporal coordinate and three spatial coordinates, respectively. The temporal coordinate, γ_0, shall remain unchanged in all cases. For cylindrical coordinates, we also retain the longitudinal axis, γ_3. The transverse axes, γ_1 and γ_2, shall be replaced by their cylindrical counterparts, ρ_1 and ϕ_2, as shown in Figure D.1(b). They have the same normalization as all other spatial axes in the $(+---)$ metric,

$$\rho_1^2 = \phi_2^2 = -1 \tag{D.60}$$

The polar bivector planes associated with extending these axes through time are

$$\rho_1 \wedge \gamma_0 = \rho_1\gamma_0 = \rho_{10} \tag{D.61a}$$

$$\phi_2 \wedge \gamma_0 = \phi_2\gamma_0 = \phi_{20} \tag{D.61b}$$

Note that the ordering was chosen so that the resulting coordinate system would be right-handed,

$$\rho_{10} \times \phi_{20} = \gamma_{30} = \sigma_3 \tag{D.62}$$

Axial bivectors can also be defined,

$$i\sigma_3 = i\left(\rho_{10} \times \phi_{20}\right) = \tfrac{1}{2}\left(\rho_{10}\phi_{20} - \phi_{20}\rho_{10}\right) = \rho_{10}\phi_{20} = \phi_2\rho_1 \tag{D.63a}$$

$$i\phi_{20} = i\phi_2\gamma_0 = i\left(-i\sigma_3\rho_1\right)\gamma_0 = \gamma_3\gamma_0\rho_1\gamma_0 = \rho_1\gamma_3 \tag{D.63b}$$

$$i\rho_{10} = -i\rho_1\gamma_3\gamma_3\gamma_0 = \phi_{20}\gamma_3\gamma_0 = \gamma_3\phi_2 \tag{D.63c}$$

Differential forms in cylindrical coordinates may be derived using a spacetime split that expresses its result in terms of the del operator ($\boldsymbol{\nabla}$), then substituting the appropriate forms for that operator from three-dimensional vector calculus, most of which were given in Appendix B. For example, the spacetime split of the divergence and curl of a bivector field are

$$\left(\Box \cdot \left(\mathbf{a} + i\mathbf{b}\right)\right)\gamma_0 = \left(\boldsymbol{\nabla} \cdot \mathbf{a}\right) - \partial_0\mathbf{a} + \boldsymbol{\nabla} \times \mathbf{b} \tag{D.64a}$$

$$\left(\Box \wedge \left(\mathbf{a} + i\mathbf{b}\right)\right)\gamma_0 i = \left(\boldsymbol{\nabla} \cdot \mathbf{b}\right) - \partial_0\mathbf{b} - \boldsymbol{\nabla} \times \mathbf{a} \tag{D.64b}$$

These can then be expanded using the vector calculus forms of the del operator in cylindrical coordinates given in (B.15b) and (B.20). These expressions are quite lengthy in general, and will not be repeated here as little simplification is possible without additional constraints.

Also of interest is the spacetime gradient of a scalar, which in cylindrical coordinates may be written as

$$\Box\psi = \gamma_0\left(\partial_0 + \boldsymbol{\nabla}\right)\psi = \gamma_0\left(\frac{1}{c}\frac{\partial\psi}{\partial t} + \frac{\partial\psi}{\partial\rho}\rho_{10} + \frac{1}{\rho}\frac{\partial\psi}{\partial\phi}\phi_{20} + \frac{\partial\psi}{\partial z}\sigma_3\right) \tag{D.65a}$$

$$= \frac{1}{c}\frac{\partial\psi}{\partial t}\gamma_0 - \left(\frac{\partial\psi}{\partial\rho}\rho_1 + \frac{1}{\rho}\frac{\partial\psi}{\partial\phi}\phi_2 + \frac{\partial\psi}{\partial z}\gamma_3\right) \tag{D.65b}$$

and the spacetime Laplacian, which is

$$\Box^2\psi = \left(\partial_0^2 - \boldsymbol{\nabla}^2\right)\psi = \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} - \left(\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\psi}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2\psi}{\partial\phi^2} + \frac{\partial^2\psi}{\partial z^2}\right) \tag{D.66}$$

Other differential forms may be derived as needed in a similar manner.

D.7.2 Spherical Coordinates

Basis vectors for spherical coordinates are shown in Figure D.1(c). In this case, all three spatial vectors are deprecated in favor of r_1, θ_2, and ϕ_3, where

$$r_1^2 = \theta_2^2 = \phi_3^2 = -1 \tag{D.67}$$

The temporal axis, γ_0, is retained. The polar bivectors follow the usual pattern,

$$r_1 \wedge \gamma_0 = r_1\gamma_0 = r_{10} \tag{D.68a}$$

$$\theta_2 \wedge \gamma_0 = \theta_2\gamma_0 = \theta_{20} \tag{D.68b}$$

$$\phi_3 \wedge \gamma_0 = \phi_3\gamma_0 = \phi_{30} \tag{D.68c}$$

Note once again that the coordinate system is right-handed as drawn in the figure,

$$r_{10} \times \theta_{20} = \phi_{30} \tag{D.69}$$

This allows us to derive the axial bivectors,

$$i\phi_{30} = \tfrac{1}{2}\left(r_{10}\theta_{20} - \theta_{20}r_{10}\right) = r_{10}\theta_{20} = \theta_2 r_1 \tag{D.70a}$$

$$i\theta_{20} = i\left(-i\phi_{30}r_1\right)\gamma_0 = \phi_{30}r_1\gamma_0 = r_1\phi_3 \tag{D.70b}$$

$$ir_{10} = -ir_1\phi_3\phi_3\gamma_0 = \theta_{20}\phi_3\gamma_0 = \phi_3\theta_2 \tag{D.70c}$$

Differential forms of the divergence and curl in spherical coordinates can be derived using the same splits found in (D.64), supplemented with the formulae from vector calculus given in (B.15c) and (B.21). The gradient is given by

$$\Box\psi = \frac{1}{c}\frac{\partial\psi}{\partial t}\gamma_0 - \left(\frac{\partial\psi}{\partial r}r_1 + \frac{1}{r}\frac{\partial\psi}{\partial\theta}\theta_2 + \frac{1}{r\sin\theta}\frac{\partial\psi}{\partial\phi}\phi_3\right) \tag{D.71}$$

while the Laplacian is

$$\Box^2\psi = \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} - \left(\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2}\right) \quad \text{(D.72)}$$

D.8 IDENTITIES

D.8.1 Commutivity

$$\psi\mathcal{M} = \mathcal{M}\psi \quad \text{(D.73)}$$

$$iK = (-1)^k Ki \quad \text{(D.74)}$$

$$a \cdot K = (-1)^{k+1} K \cdot a \quad \text{(D.75)}$$

$$K \wedge L = (-1)^{kl} L \wedge K \quad \text{(D.76)}$$

$$F \cdot K = \begin{cases} -K \cdot F & k = 1 \\ K \cdot F & \text{otherwise} \end{cases} \quad \text{(where } f = 2) \quad \text{(D.77)}$$

$$[\mathcal{M}, \mathcal{N}] = -[\mathcal{N}, \mathcal{M}] \quad \text{(D.78)}$$

$$\mathcal{M}_{even} \cdot \mathcal{N}_{even} = \mathcal{N}_{even} \cdot \mathcal{M}_{even} \quad \text{(D.79)}$$

$$\mathcal{M}_{even} \wedge \mathcal{N}_{even} = \mathcal{N}_{even} \wedge \mathcal{M}_{even} \quad \text{(D.80)}$$

$$\mathcal{M}_{even} e^{i\varphi} = e^{i\varphi}\mathcal{M}_{even} \quad \text{(D.81)}$$

$$\mathcal{M}_{odd} e^{i\varphi} = e^{-i\varphi}\mathcal{M}_{odd} \quad \text{(D.82)}$$

D.8.2 Grade Projection

$$\langle \mathcal{M}\mathcal{N} \rangle = \langle \mathcal{N}\mathcal{M} \rangle \quad \text{(D.83)}$$

$$i\langle \mathcal{M} \rangle_k = \langle i\mathcal{M} \rangle_{4-k} \quad \text{(D.84)}$$

$$\langle [F, \mathcal{M}] \rangle_k = [F, \langle \mathcal{M} \rangle_k] \quad \text{for } f = 2 \quad \text{(D.85)}$$

D.8.3 Geometric Projection

$$(\mathcal{M}+\mathcal{N})_K = \mathcal{M}_K + \mathcal{N}_K \tag{D.86}$$

$$(\mathcal{M}+\mathcal{N})_{\cancel{K}} = \mathcal{M}_{\cancel{K}} + \mathcal{N}_{\cancel{K}} \tag{D.87}$$

$$a_{\cancel{K}} = a_{iK} \tag{D.88}$$

$$K_{\cancel{a}} = K_{ia} \quad \text{for} \quad 1 \le k \le 3 \tag{D.89}$$

$$\mathcal{M} = \mathcal{M}_a + \mathcal{M}_{\cancel{a}} \tag{D.90}$$

$$a = a_K + a_{\cancel{K}} = {}_K a + {}_{\cancel{K}} a \tag{D.91}$$

$$a_b \cdot \mathcal{M}_b = a_b \cdot \mathcal{M} \tag{D.92}$$

$$a_b \cdot \mathcal{M}_{\cancel{b}} = 0 \tag{D.93}$$

$$a_{\cancel{b}} \cdot \mathcal{M}_b = (a \cdot \mathcal{M})_b \tag{D.94}$$

$$a_{\cancel{b}} \cdot \mathcal{M}_{\cancel{b}} = a \cdot \mathcal{M}_{\cancel{b}} \tag{D.95}$$

$$a_b \wedge \mathcal{M}_b = 0 \tag{D.96}$$

$$a_b \wedge \mathcal{M}_{\cancel{b}} = a_b \wedge \mathcal{M} \tag{D.97}$$

$$a_{\cancel{b}} \wedge \mathcal{M}_b = a \wedge \mathcal{M}_b \tag{D.98}$$

$$a_{\cancel{b}} \wedge \mathcal{M}_{\cancel{b}} = (a \wedge \mathcal{M})_{\cancel{b}} \tag{D.99}$$

$$a_b M_b = a_b \cdot \mathcal{M} \tag{D.100}$$

$$a_b M_{\cancel{b}} = a_b \wedge \mathcal{M} \tag{D.101}$$

$$a_b \mathcal{M} = a_b \cdot \mathcal{M}_b + a_b \wedge \mathcal{M}_{\cancel{b}} \tag{D.102}$$

$$a_{KL} = a_K + a_L \quad \text{if} \quad KL = K \wedge L \tag{D.103}$$

$$(\mathcal{M}_a)_b = (\mathcal{M}_b)_a \quad \text{if} \quad a \cdot b = 0 \tag{D.104}$$

$$(\mathcal{M}_a)_{\cancel{b}} = (\mathcal{M}_{\cancel{b}})_a \quad \text{if} \quad a \cdot b = 0 \tag{D.105}$$

$$(\mathcal{M}_{\cancel{a}})_{\cancel{b}} = (\mathcal{M}_{\cancel{b}})_{\cancel{a}} \quad \text{if} \quad a \cdot b = 0 \tag{D.106}$$

$$(\mathcal{M}_K)_K = \mathcal{M}_K \tag{D.107}$$

$$(\mathcal{M}_{\cancel{K}})_{\cancel{K}} = \mathcal{M}_{\cancel{K}} \quad \text{if} \quad k \neq 2 \tag{D.108}$$

$$(\mathcal{M}_K)_{\cancel{K}} = 0 \quad \text{if} \quad k \neq 2 \tag{D.109}$$

$$(\mathcal{M}_{\cancel{K}})_K = 0 \tag{D.110}$$

$$\mathcal{M}_K = {}_K\mathcal{M} \quad \text{if} \quad K \wedge K = 0 \tag{D.111}$$

$$\mathcal{M}_{\cancel{K}} = {}_{\cancel{K}}\mathcal{M} \quad \text{if} \quad K \wedge K = 0 \tag{D.112}$$

D.8.4 Reversion

$$(\mathcal{M}\mathcal{N})\tilde{} = \widetilde{\mathcal{N}}\widetilde{\mathcal{M}} \tag{D.113a}$$

$$(\mathcal{M}\cdot\mathcal{N})\tilde{} = \widetilde{\mathcal{N}}\cdot\widetilde{\mathcal{M}} \tag{D.113b}$$

$$(\mathcal{M}\wedge\mathcal{N})\tilde{} = \widetilde{\mathcal{N}}\wedge\widetilde{\mathcal{M}} \tag{D.113c}$$

$$[\mathcal{M},\mathcal{N}]\tilde{} = \left[\widetilde{\mathcal{N}},\widetilde{\mathcal{M}}\right] \tag{D.113d}$$

D.8.5 Product Duality

$$(a\cdot\mathcal{M})i = a\wedge(\mathcal{M}i) \tag{D.114}$$

$$(a\wedge\mathcal{M})i = a\cdot(\mathcal{M}i) \tag{D.115}$$

$$i(\mathcal{M}\cdot a) = (i\mathcal{M})\wedge a \tag{D.116}$$

$$i(\mathcal{M}\wedge a) = (i\mathcal{M})\cdot a \tag{D.117}$$

$$(F\cdot G)i = F\wedge(Gi) \quad f=g=2 \tag{D.118}$$

$$(F\wedge G)i = F\cdot(Gi) \quad f=g=2 \tag{D.119}$$

D.8.6 Binary Products

$$a\mathcal{M} = a\cdot\mathcal{M} + a\wedge\mathcal{M} \tag{D.120}$$

$$\mathcal{M}a = \mathcal{M}\cdot a + \mathcal{M}\wedge a \tag{D.121}$$

$$a\cdot K = \tfrac{1}{2}\left(aK + (-1)^{k+1}Ka\right) \tag{D.122}$$

$$a\wedge K = \tfrac{1}{2}\left(aK + (-1)^{k}Ka\right) \tag{D.123}$$

$$\{F,G\} = F\cdot G + F\wedge G \quad \text{when } f=g=2 \tag{D.124}$$

$$FG = \{F,G\} + [F,G] \quad \text{when } f=g=2 \tag{D.125}$$

$$K^2 = K\cdot K + K\wedge K \tag{D.126}$$

$$K\cdot L = \langle KL\rangle_{|k-l|} \quad \text{for } k>0 \text{ and } l>0 \tag{D.127}$$

$$K\wedge L = \langle KL\rangle_{k+l} \quad \text{for } k+l\le 4 \tag{D.128}$$

$$[\mathcal{M},\mathcal{N}] = \tfrac{1}{2}(\mathcal{M}\mathcal{N} - \mathcal{N}\mathcal{M}) \tag{D.129}$$

$$\{\mathcal{M},\mathcal{N}\} = \tfrac{1}{2}(\mathcal{M}\mathcal{N} + \mathcal{N}\mathcal{M}) \tag{D.130}$$

$$[a, K] = \begin{cases} a \cdot K & k \text{ is even} \\ a \wedge K & k \text{ is odd} \end{cases} \tag{D.131}$$

$$\{a, K\} = \begin{cases} a \wedge K & k \text{ is even} \\ a \cdot K & k \text{ is odd} \end{cases} \tag{D.132}$$

D.8.7 Trinary Products

$$a \cdot (bc) = a \cdot (b \wedge c) = (a \cdot b)c - (a \cdot c)b \tag{D.133}$$

$$a \wedge (bc) = a(b \cdot c) + a \wedge b \wedge c \tag{D.134}$$

$$a(b \wedge c) = (a \cdot b)c - (a \cdot c)b + a \wedge b \wedge c \tag{D.135}$$

$$abc = (a \cdot b)c - (a \cdot c)b + (b \cdot c)a + a \wedge b \wedge c \tag{D.136}$$

$$a \cdot (F \wedge K) = (a \cdot F) \wedge K + F \wedge (a \cdot K) \quad \text{for } f = 2 \tag{D.137}$$

$$(L \wedge a) \cdot K = L \cdot (a \cdot K) \quad \text{for} \quad k \geq 2 \text{ and } 1 \leq l < k \tag{D.138}$$

$$K \cdot (a \wedge L) = (K \cdot a) \cdot L \quad \text{for} \quad k \geq 2 \text{ and } 1 \leq l < k \tag{D.139}$$

$$a \cdot (b \cdot K) = -b \cdot (a \cdot K) = (K \cdot a) \cdot b \tag{D.140}$$

$$a \cdot (bK) = (a \cdot b)K - b(a \cdot K) \tag{D.141}$$

$$a \cdot (b \wedge K) = (a \cdot b)K - b \wedge (a \cdot K) \tag{D.142}$$

$$a \wedge (b \cdot K) = (a \cdot b)K - b \cdot (a \wedge K) \tag{D.143}$$

$$a \cdot (K \wedge L) = (a \cdot K) \wedge L + (-1)^k K \wedge (a \cdot L) \tag{D.144}$$

$$[\mathcal{M},\mathcal{N}\mathcal{O}] = [\mathcal{M},\mathcal{N}]\,\mathcal{O} + \mathcal{N}\,[\mathcal{M},\mathcal{O}] \tag{D.145}$$

$$[\mathcal{M},[\mathcal{N},\mathcal{O}]] = [[\mathcal{M},\mathcal{N}]\,,\mathcal{O}] + [\mathcal{N},[\mathcal{M},\mathcal{O}]] \tag{D.146}$$

D.8.8 Quaternary and Higher-Order Products

$$(a \wedge b) \cdot (c \wedge d) = (a \cdot d)(b \cdot c) - (a \cdot c)(b \cdot d) \tag{D.147}$$

$$a \cdot (b \wedge c \wedge d) = (a \cdot b)(c \wedge d) - (a \cdot c)(b \wedge d) + (a \cdot d)(b \wedge c) \tag{D.148}$$

D.8.9 Degenerate Forms

$$(\mathcal{M} \cdot a) \cdot a = 0 \tag{D.149}$$

$$\mathcal{M} \wedge a \wedge a = 0 \tag{D.150}$$

$$(\mathcal{M} \cdot a) \wedge a = (\mathcal{M} \cdot a)a = (\mathcal{M}a) \wedge a \tag{D.151}$$

$$(\mathcal{M} \wedge a) \cdot a = (\mathcal{M} \wedge a)a = (\mathcal{M}a) \cdot a \tag{D.152}$$

$$K_i = K_{\underline{1}} = K \quad k > 0 \tag{D.153}$$

$$K_{\underline{i}} = K_1 = 0 \quad k > 0 \tag{D.154}$$

D.8.10 Differential Identities

Because the spacetime gradient, $\square$, behaves like a vector, most of the product identities in the previous sections can be converted into a differential form so long as the product rule for differentiation is observed,

$$\square(\mathcal{M}\mathcal{N}) = \dot{\square}\dot{\mathcal{M}}\mathcal{N} + \dot{\square}\mathcal{M}\dot{\mathcal{N}} \tag{D.155}$$

where the overdots indicate which element is differentiated in each term (we assume in this section that each object potentially depends on the four-position, x, and is thus subject to differentiation). To list some examples,

$$\square\mathcal{M} = \square \cdot \mathcal{M} + \square \wedge \mathcal{M} \tag{D.156}$$

$$\square \cdot (ab) = \square \cdot (a \wedge b) = (a \cdot \square)b - (b \cdot \square)a + b(\square \cdot a) - a(\square \cdot b) \tag{D.157}$$

$$(\square \cdot K)i = \square \wedge (Ki) \tag{D.158}$$

$$(\square \wedge K)i = \square \cdot (Ki) \tag{D.159}$$

$$(a \wedge \square) \cdot K = a \cdot (\square \cdot K) \text{ for } k \geq 2 \tag{D.160}$$

$$K \cdot (\square \wedge b) = (K \cdot \square) \cdot b \text{ for } k \geq 2 \tag{D.161}$$

$$a \cdot (\square \wedge K) = (a \cdot \square)K - \dot{\square} \wedge (a \cdot \dot{K}) \tag{D.162}$$

$$a \wedge (\square \cdot K) = (a \cdot \square)K - \dot{\square} \cdot (a \wedge \dot{K}) \tag{D.163}$$

$$\square \cdot (\square \wedge K) = \square \cdot (\square K) = \square(\square \wedge K) \tag{D.164}$$

$$\square \wedge (\square \cdot K) = \square \wedge (\square K) = \square(\square \cdot K) \tag{D.165}$$

$$\square^2 K = \square(\square \cdot K) + \square \cdot (\square K) \tag{D.166}$$

$$\Box^2 K = \Box(\Box \wedge K) + \Box \wedge (\Box K) \tag{D.167}$$

$$\Box \cdot (\Box \cdot \mathcal{M}) = 0 \tag{D.168}$$

$$\Box \wedge \Box \wedge \mathcal{M} = 0 \tag{D.169}$$

$$\Box_{KL} = \Box_K + \Box_L \quad \text{if} \quad KL = K \wedge L \tag{D.170}$$

References

[1] J. Dressel, K. Y. Bliokh, and F. Nori, "Spacetime algebra as a powerful tool for electromagnetism," *Physics Reports*, vol. 589, pp. 1–71, 2015.

[2] C. Doran, A. Lasenby, S. Gull, S. Somaroo, and A. Challinor, "Spacetime algebra and electron physics," *Advances in Imaging and Electron Physics*, vol. 95, pp. 271–386, 1996.

[3] D. Hestenes. (1998) Spacetime calculus. http://geocalc.clas.asu.edu/pdf/SpaceTimeCalc.pdf.

[4] D. Hestenes, "A unified language for mathematics and physics," *Clifford Algebras and Their Applications in Mathematical Physics*, pp. 1–23, 1986.

[5] J. Lasenby, A. N. Lasenby, and C. J. L. Doran, "A unified mathematical language for physics and engineering in the 21st century," *Philosophical Transactions of the Royal Society of London A*, vol. 358, no. 1765, pp. 21–39, January 2000.

[6] Wikipedia. (2024, January) Spacetime algebra. https://en.wikipedia.org/wiki/Spacetime_algebra.

[7] A. Arsenovic, "Applications of conformal geometric algebra to transmission line theory," *IEEE Access*, vol. 5, pp. 19 920–19 941, July 2017.

[8] D. Hestenes, "Multivector calculus," *Journal of Mathematical Analysis and Applications*, vol. 24, no. 2, pp. 313–325, 1968.

[9] D. Hestenes, "Multivector functions," *Journal of Mathematical Analysis and Applications*, vol. 24, no. 3, pp. 467–473, 1968.

[10] D. Hestenes, "Differential forms in geometric calculus," *Proceedings of the Third Conference on Clifford Algebras and their Applications in Mathematical Physics*, vol. 55, pp. 269–285, 1993.

Appendix E

Metric Signatures

An unfortunate reality of this subject matter is that different authors use different conventions regarding the metric signature and the ordering of spatial and temporal components of four-vectors and tensors. This leads to numerous subtle discrepancies in the resulting equations, usually in the form of missing minus signs or alternate indices of summation.

To facilitate the comparison of results from references using different conventions, I present the following tables that reveal the key results of this book in several of the most common forms.

E.1 TENSOR ANALYSIS

E.1.1 First and Second-Rank Tensors

The most critical four-vectors, or rank-1 tensors, are listed in both covariant and contravariant forms in Table E.1 using four different metric signatures. Second-rank tensors are given in Table E.2. Recall that

$$[\mathbf{A}]_\times = \begin{pmatrix} 0 & -A_z & A_y \\ A_z & 0 & -A_x \\ -A_y & A_x & 0 \end{pmatrix} \tag{E.1}$$

The metric signature listed first in each table is the one used throughout this book. The outermost signature columns having mostly minuses follow the timelike or *Landau-Lifshitz* convention, which is popular in particle physics. The inner two

Table E.1

Four-Vector Forms for Various Metric Signatures

Name	Symbol	(+ − − −)	(− + + +)	(+ + + −)	(− − − +)
		Contravariant Forms			
Four-position	X^μ	$(ct, \mathbf{r})$	$(ct, \mathbf{r})$	$(\mathbf{r}, ct)$	$(\mathbf{r}, ct)$
Four-velocity	U^μ	$(c, \mathbf{v})\gamma$	$(c, \mathbf{v})\gamma$	$(\mathbf{v}, c)\gamma$	$(\mathbf{v}, c)\gamma$
Four-momentum	P^μ	$(\frac{E}{c}, \mathbf{p})$	$(\frac{E}{c}, \mathbf{p})$	$(\mathbf{p}, \frac{E}{c})$	$(\mathbf{p}, \frac{E}{c})$
Four-force	F^μ	$(\frac{\dot{E}}{c}, \dot{\mathbf{p}})\gamma$	$(\frac{\dot{E}}{c}, \dot{\mathbf{p}})\gamma$	$(\dot{\mathbf{p}}, \frac{\dot{E}}{c})\gamma$	$(\dot{\mathbf{p}}, \frac{\dot{E}}{c})\gamma$
Four-wavevector	K^μ	$(\frac{\omega}{c}, \mathbf{k})$	$(\frac{\omega}{c}, \mathbf{k})$	$(\mathbf{k}, \frac{\omega}{c})$	$(\mathbf{k}, \frac{\omega}{c})$
Four-current	J^μ	$(c\rho, \mathbf{J})$	$(c\rho, \mathbf{J})$	$(\mathbf{J}, c\rho)$	$(\mathbf{J}, c\rho)$
Four-potential	A^μ	$(\frac{\varphi}{c}, \mathbf{A})$	$(\frac{\varphi}{c}, \mathbf{A})$	$(\mathbf{A}, \frac{\varphi}{c})$	$(\mathbf{A}, \frac{\varphi}{c})$
Four-gradient	∂^μ	$(\frac{1}{c}\frac{\partial}{\partial t}, -\boldsymbol{\nabla})$	$(-\frac{1}{c}\frac{\partial}{\partial t}, \boldsymbol{\nabla})$	$(\boldsymbol{\nabla}, -\frac{1}{c}\frac{\partial}{\partial t})$	$(-\boldsymbol{\nabla}, \frac{1}{c}\frac{\partial}{\partial t})$
		Covariant Forms			
Four-position	X_μ	$(ct, -\mathbf{r})$	$(-ct, \mathbf{r})$	$(\mathbf{r}, -ct)$	$(-\mathbf{r}, ct)$
Four-velocity	U_μ	$(c, -\mathbf{v})\gamma$	$(-c, \mathbf{v})\gamma$	$(\mathbf{v}, -c)\gamma$	$(-\mathbf{v}, c)\gamma$
Four-momentum	P_μ	$(\frac{E}{c}, -\mathbf{p})$	$(-\frac{E}{c}, \mathbf{p})$	$(\mathbf{p}, -\frac{E}{c})$	$(-\mathbf{p}, \frac{E}{c})$
Four-force	F_μ	$(\frac{\dot{E}}{c}, -\dot{\mathbf{p}})\gamma$	$(-\frac{\dot{E}}{c}, \dot{\mathbf{p}})\gamma$	$(\dot{\mathbf{p}}, -\frac{\dot{E}}{c})\gamma$	$(-\dot{\mathbf{p}}, \frac{\dot{E}}{c})\gamma$
Four-wavevector	K_μ	$(\frac{\omega}{c}, -\mathbf{k})$	$(-\frac{\omega}{c}, \mathbf{k})$	$(\mathbf{k}, -\frac{\omega}{c})$	$(-\mathbf{k}, \frac{\omega}{c})$
Four-current	J_μ	$(c\rho, -\mathbf{J})$	$(-c\rho, \mathbf{J})$	$(\mathbf{J}, -c\rho)$	$(-\mathbf{J}, c\rho)$
Four-potential	A_μ	$(\frac{\varphi}{c}, -\mathbf{A})$	$(-\frac{\varphi}{c}, \mathbf{A})$	$(\mathbf{A}, -\frac{\varphi}{c})$	$(-\mathbf{A}, \frac{\varphi}{c})$
Four-gradient	∂_μ	$(\frac{1}{c}\frac{\partial}{\partial t}, \boldsymbol{\nabla})$	$(\frac{1}{c}\frac{\partial}{\partial t}, \boldsymbol{\nabla})$	$(\boldsymbol{\nabla}, \frac{1}{c}\frac{\partial}{\partial t})$	$(\boldsymbol{\nabla}, \frac{1}{c}\frac{\partial}{\partial t})$

Table E.2

Tensor Forms for Various Metric Signatures

Symbol	(+ − − −)	(− + + +)	(+ + + −)	(− − − +)
Contravariant Forms				
$F^{\mu\nu}$	$\begin{pmatrix} 0 & -\frac{1}{c}\mathbf{E}^T \\ \frac{1}{c}\mathbf{E} & [\mathbf{B}]_\times \end{pmatrix}$	$\begin{pmatrix} 0 & \frac{1}{c}\mathbf{E}^T \\ -\frac{1}{c}\mathbf{E} & -[\mathbf{B}]_\times \end{pmatrix}$	$\begin{pmatrix} -[\mathbf{B}]_\times & -\frac{1}{c}\mathbf{E} \\ \frac{1}{c}\mathbf{E}^T & 0 \end{pmatrix}$	$\begin{pmatrix} [\mathbf{B}]_\times & \frac{1}{c}\mathbf{E} \\ -\frac{1}{c}\mathbf{E}^T & 0 \end{pmatrix}$
$G^{\mu\nu}$	$\begin{pmatrix} 0 & -\mathbf{B}^T \\ \mathbf{B} & -\frac{1}{c}[\mathbf{E}]_\times \end{pmatrix}$	$\begin{pmatrix} 0 & \mathbf{B}^T \\ -\mathbf{B} & \frac{1}{c}[\mathbf{E}]_\times \end{pmatrix}$	$\begin{pmatrix} \frac{1}{c}[\mathbf{E}]_\times & -\mathbf{B} \\ \mathbf{B}^T & 0 \end{pmatrix}$	$\begin{pmatrix} -\frac{1}{c}[\mathbf{E}]_\times & \mathbf{B} \\ -\mathbf{B}^T & 0 \end{pmatrix}$
$\mathcal{D}^{\mu\nu}$	$\begin{pmatrix} 0 & -c\mathbf{D}^T \\ c\mathbf{D} & [\mathbf{H}]_\times \end{pmatrix}$	$\begin{pmatrix} 0 & c\mathbf{D}^T \\ -c\mathbf{D} & -[\mathbf{H}]_\times \end{pmatrix}$	$\begin{pmatrix} -[\mathbf{H}]_\times & -c\mathbf{D} \\ c\mathbf{D}^T & 0 \end{pmatrix}$	$\begin{pmatrix} [\mathbf{H}]_\times & c\mathbf{D} \\ -c\mathbf{D}^T & 0 \end{pmatrix}$
$\star\mathcal{D}^{\mu\nu}$	$\begin{pmatrix} 0 & -\mathbf{H}^T \\ \mathbf{H} & -c[\mathbf{D}]_\times \end{pmatrix}$	$\begin{pmatrix} 0 & \mathbf{H}^T \\ -\mathbf{H} & c[\mathbf{D}]_\times \end{pmatrix}$	$\begin{pmatrix} c[\mathbf{D}]_\times & -\mathbf{H} \\ \mathbf{H}^T & 0 \end{pmatrix}$	$\begin{pmatrix} -c[\mathbf{D}]_\times & \mathbf{H} \\ -\mathbf{H}^T & 0 \end{pmatrix}$
$\mathcal{M}^{\mu\nu}$	$\begin{pmatrix} 0 & c\mathbf{P}^T \\ -c\mathbf{P} & [\mathbf{M}]_\times \end{pmatrix}$	$\begin{pmatrix} 0 & -c\mathbf{P}^T \\ c\mathbf{P} & -[\mathbf{M}]_\times \end{pmatrix}$	$\begin{pmatrix} -[\mathbf{M}]_\times & c\mathbf{P} \\ -c\mathbf{P}^T & 0 \end{pmatrix}$	$\begin{pmatrix} [\mathbf{M}]_\times & -c\mathbf{P} \\ c\mathbf{P}^T & 0 \end{pmatrix}$
$T^{\mu\nu}$	$\begin{pmatrix} u & \frac{1}{c}\mathbf{S}^T \\ \frac{1}{c}\mathbf{S} & -\boldsymbol{\sigma} \end{pmatrix}$	$\begin{pmatrix} u & \frac{1}{c}\mathbf{S}^T \\ \frac{1}{c}\mathbf{S} & -\boldsymbol{\sigma} \end{pmatrix}$	$\begin{pmatrix} -\boldsymbol{\sigma} & \frac{1}{c}\mathbf{S} \\ \frac{1}{c}\mathbf{S}^T & u \end{pmatrix}$	$\begin{pmatrix} -\boldsymbol{\sigma} & \frac{1}{c}\mathbf{S} \\ \frac{1}{c}\mathbf{S}^T & u \end{pmatrix}$
Covariant Forms				
$F_{\mu\nu}$	$\begin{pmatrix} 0 & \frac{1}{c}\mathbf{E}^T \\ -\frac{1}{c}\mathbf{E} & [\mathbf{B}]_\times \end{pmatrix}$	$\begin{pmatrix} 0 & -\frac{1}{c}\mathbf{E}^T \\ \frac{1}{c}\mathbf{E} & -[\mathbf{B}]_\times \end{pmatrix}$	$\begin{pmatrix} -[\mathbf{B}]_\times & \frac{1}{c}\mathbf{E} \\ -\frac{1}{c}\mathbf{E}^T & 0 \end{pmatrix}$	$\begin{pmatrix} [\mathbf{B}]_\times & -\frac{1}{c}\mathbf{E} \\ \frac{1}{c}\mathbf{E}^T & 0 \end{pmatrix}$
$G_{\mu\nu}$	$\begin{pmatrix} 0 & \mathbf{B}^T \\ -\mathbf{B} & -\frac{1}{c}[\mathbf{E}]_\times \end{pmatrix}$	$\begin{pmatrix} 0 & -\mathbf{B}^T \\ \mathbf{B} & \frac{1}{c}[\mathbf{E}]_\times \end{pmatrix}$	$\begin{pmatrix} \frac{1}{c}[\mathbf{E}]_\times & \mathbf{B} \\ -\mathbf{B}^T & 0 \end{pmatrix}$	$\begin{pmatrix} -\frac{1}{c}[\mathbf{E}]_\times & -\mathbf{B} \\ \mathbf{B}^T & 0 \end{pmatrix}$
$\mathcal{D}_{\mu\nu}$	$\begin{pmatrix} 0 & c\mathbf{D}^T \\ -c\mathbf{D} & [\mathbf{H}]_\times \end{pmatrix}$	$\begin{pmatrix} 0 & -c\mathbf{D}^T \\ c\mathbf{D} & -[\mathbf{H}]_\times \end{pmatrix}$	$\begin{pmatrix} -[\mathbf{H}]_\times & c\mathbf{D} \\ -c\mathbf{D}^T & 0 \end{pmatrix}$	$\begin{pmatrix} [\mathbf{H}]_\times & -c\mathbf{D} \\ c\mathbf{D}^T & 0 \end{pmatrix}$
$\star\mathcal{D}_{\mu\nu}$	$\begin{pmatrix} 0 & \mathbf{H}^T \\ -\mathbf{H} & -c[\mathbf{D}]_\times \end{pmatrix}$	$\begin{pmatrix} 0 & -\mathbf{H}^T \\ \mathbf{H} & c[\mathbf{D}]_\times \end{pmatrix}$	$\begin{pmatrix} c[\mathbf{D}]_\times & \mathbf{H} \\ -\mathbf{H}^T & 0 \end{pmatrix}$	$\begin{pmatrix} -c[\mathbf{D}]_\times & -\mathbf{H} \\ \mathbf{H}^T & 0 \end{pmatrix}$
$\mathcal{M}_{\mu\nu}$	$\begin{pmatrix} 0 & -c\mathbf{P}^T \\ c\mathbf{P} & [\mathbf{M}]_\times \end{pmatrix}$	$\begin{pmatrix} 0 & c\mathbf{P}^T \\ -c\mathbf{P} & -[\mathbf{M}]_\times \end{pmatrix}$	$\begin{pmatrix} -[\mathbf{M}]_\times & -c\mathbf{P} \\ c\mathbf{P}^T & 0 \end{pmatrix}$	$\begin{pmatrix} [\mathbf{M}]_\times & c\mathbf{P} \\ -c\mathbf{P}^T & 0 \end{pmatrix}$

Table E.3
Key Tensor Equations for Various Metric Signatures

Equation	(+ − − −) or (− − − +)	(− + + +) or (+ + + −)
Four-vector transf.	$X'^{\mu} = \Lambda^{\mu}{}_{\nu} X^{\nu}$	$X'^{\mu} = \Lambda^{\mu}{}_{\nu} X^{\nu}$
Tensor transf.	$F'^{\mu\nu} = \Lambda^{\mu}{}_{\sigma}\Lambda^{\nu}{}_{\tau} F^{\sigma\tau}$	$F'^{\mu\nu} = \Lambda^{\mu}{}_{\sigma}\Lambda^{\nu}{}_{\tau} F^{\sigma\tau}$
Proper time	$(c\tau)^2 = X_\mu X^\mu$	$(c\tau)^2 = -X_\mu X^\mu$
Spacetime interval	$s^2 = -X_\mu X^\mu$	$s^2 = X_\mu X^\mu$
Four-velocity norm.	$U_\mu U^\mu = c^2$	$U_\mu U^\mu = -c^2$
Wavevector norm.	$K_\mu K^\mu = -\left(\frac{\omega_0}{c}\right)^2$	$K_\mu K^\mu = \left(\frac{\omega_0}{c}\right)^2$
Lorentz phase	$K_\mu X^\mu = \Phi$	$K_\mu X^\mu = -\Phi$
Charge continuity	$\partial_\mu J^\mu = 0$	$\partial_\mu J^\mu = 0$
d'Alembertian	$\Box^2 = \partial_\mu \partial^\mu = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2$	$\Box^2 = \partial_\mu \partial^\mu = \nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}$
Lorenz gauge	$\partial_\mu A^\mu = 0$	$\partial_\mu A^\mu = 0$
Four-potential	$\partial_\mu \partial^\mu A^\nu = \mu_0 J^\nu$	$\partial_\mu \partial^\mu A^\nu = -\mu_0 J^\nu$
Faraday tensor	$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$	$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$
Stress-energy	$T^{\mu\nu} = -\frac{1}{2\mu_0}\left(F^{\mu}{}_{\sigma} F^{\nu\sigma} + G^{\mu}{}_{\sigma} G^{\nu\sigma}\right)$	$T^{\mu\nu} = \frac{1}{2\mu_0}\left(F^{\mu}{}_{\sigma} F^{\nu\sigma} + G^{\mu}{}_{\sigma} G^{\nu\sigma}\right)$
Maxwell (inhom.)	$\partial_\mu F^{\mu\nu} = \mu_0 J^\nu$	$\partial_\nu F^{\mu\nu} = \mu_0 J^\mu$
Maxwell (hom.)	$\partial_\mu G^{\mu\nu} = 0$	$\partial_\nu G^{\mu\nu} = 0$
Maxwell (macro)	$\partial_\mu \mathcal{D}^{\mu\nu} = J^\nu_{free}$	$\partial_\nu \mathcal{D}^{\mu\nu} = J^\mu_{free}$
Force density	$f^\mu = F^{\mu\nu} J_\nu = -\partial_\nu T^{\mu\nu}$	$f^\mu = F^{\mu\nu} J_\nu = -\partial_\nu T^{\mu\nu}$

columns, having mostly pluses, adopt the spacelike or *Pauli* convention, most often used in relativity.

E.1.2 Tensor Equations

As a consequence of the variations listed in Tables E.1 and E.2, the choice of metric signature affects the forms of certain tensor equations as well. A summary of the key tensor equations in both standard metric forms, whether the same or different, is shown in Table E.3. Usually the difference is a minus sign, but sometimes we contract a different set of indices to avoid the minus sign, taking advantage of the field tensor asymmetry (e.g., in the inhomogeneous Maxwell's equation).

Finally, the Hodge dual in terms of the Levi-Civita symbol depends on how that symbol is defined with respect to the index ranges. In this book, we have used the convention that Greek tensor indices range from 0 to 3, and $\epsilon^{0123} = -1$, but some authors use the range 1 to 4 (with 4 corresponding to the temporal coordinate) and $\epsilon^{1234} = 1$. Therefore,

$$G^{\mu\nu} = \star F^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\sigma\tau} F_{\sigma\tau} \begin{cases} -1 & \text{index range is } 1\ldots 4 \text{ and } \epsilon^{1234} = 1 \\ 1 & \text{index range is } 0\ldots 3 \text{ and } \epsilon^{0123} = -1 \end{cases} \tag{E.2}$$

E.2 SPACETIME ALGEBRA

Unlike the tensor formalism, spacetime algebra is intrinsically coordinate-free, so that no explicit difference in the construction of mathematical objects (four-vectors, bivectors, pseudoscalars, etc.) from their arbitrary basis components (1, γ_μ, σ_k, $i\sigma_k$, and i) as a consequence of the choice of metric is either needed or implied. For example, the spacetime gradient, four-vector potential, and Faraday bivector are always given by

$$\Box = \gamma^\mu \partial_\mu \tag{E.3a}$$

$$A = A^\mu \gamma_\mu \tag{E.3b}$$

$$F = \left(\tfrac{1}{c}E^k + iB^k\right)\sigma_k \tag{E.3c}$$

regardless of the metric signature used. Instead, the metric is manifested in the definition of products between vector components, so that

$$F = \begin{cases} \Box \wedge A & \text{for signature } (+---) \\ -\Box \wedge A & \text{for signature } (-+++) \end{cases} \tag{E.4}$$

The purely algebraic consequences of the metric signature (having no direct connection to electromagnetics) are summarized in Table E.4. All coordinate-free operational relationships and identities, such as those listed in Section D.8, are unaffected. Spacetime splits and other vector calculus expansions are given in Table E.5, while key electromagnetic laws, whether or not they are impacted by the metric signature, are listed in Table E.6 and Table E.7.

Table E.4
Coordinate-Specific Algebraic Relationships for Different Metric Signatures

Expression	(+ − − −)	(− + + +)	Expression	(+ − − −)	(− + + +)
γ_0^2	1	-1	$i\gamma_0$	γ_{321}	γ_{123}
γ_k^2	-1	1	$i\gamma_1$	γ_{032}	γ_{230}
γ^0	γ_0	$-\gamma_0$	$i\gamma_2$	γ_{301}	γ_{103}
γ^k	$-\gamma_k$	γ_k	$i\gamma_3$	γ_{210}	γ_{012}
$\sigma_j\sigma_k$	$\gamma_k\gamma_j$	$\gamma_j\gamma_k$	$i\sigma_1$	γ_{32}	γ_{32}
$\sigma_j \cdot \sigma_k$	δ_{jk}	δ_{jk}	$i\sigma_2$	γ_{13}	γ_{13}
$\sigma_k \cdot \gamma_0$	γ_k	$-\gamma_k$	$i\sigma_3$	γ_{21}	γ_{21}
$\sigma_j \cdot \gamma_k$	$\gamma_0\delta_{jk}$	$-\gamma_0\delta_{jk}$	$\sigma_1\sigma_2\sigma_3$	i	$-i$

$\mu = 0\ldots 3, \quad j,k = 1\ldots 3.$

Table E.5
Spacetime Splits and Vector Calculus Expansion for Different Metric Signatures

Split/Expansion	(+ − − −)	(− + + +)
$ct + \mathbf{r}$	$x\gamma_0$	$\gamma^0 x$
$\partial_0 + \boldsymbol{\nabla}$	$\gamma^0\Box$	$\Box\gamma_0$
$\mathbf{a}\cdot\mathbf{b}$	$\frac{1}{2}(\mathbf{ab}+\mathbf{ba})$	$\frac{1}{2}(\mathbf{ab}+\mathbf{ba})$
$\mathbf{a}\times\mathbf{b}$	$-\frac{i}{2}(\mathbf{ab}-\mathbf{ba})$	$\frac{i}{2}(\mathbf{ab}-\mathbf{ba})$
$(\mathbf{k}\cdot\mathbf{a}) + \left(k^0\mathbf{a} + \mathbf{k}\times\mathbf{b}\right)$	$((\mathbf{a}+i\mathbf{b})\cdot k)\,\gamma_0$	$\gamma^0\,(k\cdot(\mathbf{a}+i\mathbf{b}))$
$(\mathbf{k}\cdot\mathbf{b}) + \left(k^0\mathbf{b} + \mathbf{k}\times\mathbf{a}\right)$	$((\mathbf{a}-i\mathbf{b})\wedge k)\,\gamma_0 i$	$i\gamma^0\,(k\wedge(\mathbf{a}-i\mathbf{b}))$
$(\boldsymbol{\nabla}\cdot\mathbf{a}) - (\partial_0\mathbf{a} + \boldsymbol{\nabla}\times\mathbf{b})$	$(\Box\cdot(\mathbf{a}-i\mathbf{b}))\,\gamma_0$	$\gamma^0\,(\Box\cdot(\mathbf{a}-i\mathbf{b}))$
$(\boldsymbol{\nabla}\cdot\mathbf{b}) - (\partial_0\mathbf{b} + \boldsymbol{\nabla}\times\mathbf{a})$	$(\Box\wedge(\mathbf{a}+i\mathbf{b}))\,\gamma_0 i$	$-i\gamma^0\,(\Box\wedge(\mathbf{a}+i\mathbf{b}))$

Table E.6

Key Electromagnetic Laws for Different Metric Signatures

Equation	(+ − − −)	(− + + +)
Lorentz phase	$K \cdot x = \Phi$	$K \cdot x = -\Phi$
Charge continuity	$\Box \cdot J = 0$	$\Box \cdot J = 0$
Potential wave eq.	$\Box^2 A = \mu_0 J$	$\Box^2 A = -\mu_0 J$
Lorenz gauge	$\Box \cdot A = 0$	$\Box \cdot A = 0$
Faraday bivector (gauge free)	$F = \Box \wedge A$	$F = -\Box \wedge A$
Faraday bivector (Lorenz gauge)	$F = \Box A$	$F = -\Box A$
Maxwell's equation (Lorenz gauge)	$\Box F = \mu_0 J$	$\Box F = \mu_0 J$
Constitutive relation	$\mathcal{D} = c^2 \varepsilon F_U + \frac{1}{\mu} F_{\not{U}}$	$\mathcal{D} = c^2 \varepsilon F_U + \frac{1}{\mu} F_{\not{U}}$
Ohm's law	$J = \sigma F \cdot U$	$J = \sigma U \cdot F$

Table E.7

Macroscopic Electromagnetic Laws for Different Metric Signatures

Equation	(+ − − −)	(− + + +)
Charge continuity	$\blacksquare \cdot \mathcal{J} = 0$	$\blacksquare \cdot \mathcal{J} = 0$
Potential wave eq.	$\blacksquare^2 \mathcal{A} = \mathcal{J}$	$\blacksquare^2 \mathcal{A} = -\mathcal{J}$
Lorenz gauge	$\blacksquare \cdot \mathcal{A} = 0$	$\blacksquare \cdot \mathcal{A} = 0$
Faraday bivector (gauge free)	$\mathcal{F} = \blacksquare \wedge \mathcal{A}$	$\mathcal{F} = -\blacksquare \wedge \mathcal{A}$
Faraday bivector (Lorenz gauge)	$\mathcal{F} = \blacksquare \mathcal{A}$	$\mathcal{F} = -\blacksquare \mathcal{A}$
Maxwell's equation (Lorenz gauge)	$\blacksquare \mathcal{F} = \mathcal{J}$	$\blacksquare \mathcal{F} = \mathcal{J}$
Ohm's law	$\mathcal{J} = \frac{\eta \sigma}{c} \mathcal{F} \cdot U$	$\mathcal{J} = \frac{\eta \sigma}{c} \mathcal{F} \cdot U$

Appendix F

Lorentz and Lorenz

It is a virtual inevitability that any student of electromagnetics will come across the names Lorentz and Lorenz at some point in their studies, but I suspect all too few ever realize that the names refer to two different people — Hendrik Antoon Lorentz (1853–1928) and Ludvig Valentin Lorenz (1829–1891). Both individuals made significant and lasting contributions to our current understanding of the universe, deserving of recognition. Brief accounts of their works, which overlapped on more than one occasion, are given in this appendix.

F.1 HENDRIK ANTOON LORENTZ

Hendrik Lorentz [1] was a Dutch physicist who is best remembered (at least in the field of electrodynamics) for the Lorentz law describing the force imparted by the electric and magnetic fields upon a charged particle,

$$\mathbf{F} = q\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right) \tag{F.1}$$

and for first deriving the transformation equations embodying length contraction and time dilation that govern how physical quantities vary between inertial reference frames,

$$ct' = (ct - \mathbf{r} \cdot \boldsymbol{\beta})\gamma \tag{F.2}$$

$$\mathbf{r}' = \mathbf{r} + (\gamma - 1)(\mathbf{r} \cdot \mathbf{n})\mathbf{n} - \gamma ct\boldsymbol{\beta} \tag{F.3}$$

Professor to Albert Einstein, Lorentz is recognized for laying a foundation of results and principles which later became tenets of special relativity — originally known as the *Lorentz-Einstein theory* for this very reason.

Lorentz shared the Nobel Prize in physics in 1902 with Pieter Zeeman, who discovered the *Zeeman effect*, the splitting of molecular spectral lines in the presence of a magnetostatic field. Lorentz had previously suggested that atoms comprise charged particles, and furnished a satisfactory theoretical explanation for the Zeeman effect on that basis.

F.2 LUDVIG VALENTIN LORENZ

Ludvig Lorenz [2] was a Danish physicist and mathematician most often remembered for setting forth the Lorenz gauge condition

$$\frac{1}{c^2}\frac{\partial \phi}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{A} = 0 \tag{F.4}$$

which, incidentally, is a Lorentz-invariant condition (with a "t," in honor of his Dutch counterpart, Hendrik). Lesser known than Hendrik Lorentz and others (unjustly, according to some [3]), he is nonetheless an important figure in the subject matter of this book.

Lorenz worked extensively on the mathematical description of light passing through dielectric media and published a theory of light scattering in 1890, known later as *Lorenz-Mie theory*, in honor of both Lorenz and the German physicist Gustav Mie, who rediscovered it independently in 1908.

In 1869, Lorenz published a formula for the refractive index of a material in terms of its density and electric polarizability. This formula would be repeated almost a decade later by Hendrik Lorentz, earning it the name of the *Lorentz-Lorenz equation*. It is entirely equivalent to the *Clausius-Mossotti relation* which expresses the same relationship in terms of the dielectric constant in place of the refractive index.

References

[1] Wikipedia. (2024, January) Hendrik Lorentz. https://en.wikipedia.org/wiki/Hendrik_Lorentz.

[2] Wikipedia. (2023, March) Ludvig Lorenz. https://en.wikipedia.org/wiki/Ludvig_Lorenz.

[3] H. Kragh, *Ludvig Lorenz: A Nineteenth-Century Theoretical Physicist*. Denmark: The Royal Danish Academy of Sciences and Letters, Series M, Mathematica et Physica, 2018, vol. 3.

About the Author

Matthew A. Morgan received his BS in electrical engineering from the University of Virginia in 1999, and his MS and PhD in electrical engineering from the California Institute of Technology in 2001 and 2003, respectively. He has authored over 60 papers and holds 20 patents in the areas of microwave monolithic integrated circuit (MMIC) design, millimeter-wave system integration and packaging techniques, reflectionless filter development, high-speed serial communication, and ultrawideband millimeter-wave antennas.

During the summers of 1996 through 1998, he worked for Lockheed Martin Federal Systems in Manassas, Virginia, as an associate programmer, where he wrote code for acoustic signal processing, mathematical modeling, data simulation, and system performance monitoring. In 1999, he became an affiliate of NASA's Jet Propulsion Laboratory in Pasadena, California. There, he conducted research in the development of MMICs and MMIC-based receiver components for atmospheric radiometers, laboratory instrumentation, and the deep-space communication network.

In 2003, Dr. Morgan joined the Central Development Lab (CDL) of the National Radio Astronomy Observatory (NRAO) in Charlottesville, Virginia, where he now holds the position of scientist/research engineer with tenure. He was project engineer for the K-Band Focal Plane Array development on the Green Bank Telescope, and technical lead for Band 6 (211–275 GHz) cryogenic IF amplifier production, Band 6 receiver cartridge testing, and Band 3 (84–116 GHz) and Band 6 Orthomode Transducer (OMT) production for the Atacama Large Millimeter Array (ALMA). Dr. Morgan is currently the head of the CDL's Integrated Receiver Development program, and is involved in the design and development of low-noise receivers, components, and novel concepts for radio astronomy instrumentation in the centimeter-wave, millimeter-wave, and submillimeter-wave frequency ranges.

Index

D

E

F

G

H

I

J

K

L

M

Q

R

U

V

W

Y

Z

Artech House Microwave Library

Behavioral Modeling and Linearization of RF Power Amplifiers, John Wood

Chipless RFID Reader Architecture, Nemai Chandra Karmakar, Prasanna Kalansuriya, Randika Koswatta, and Rubayet E-Azim

Chipless RFID Systems Using Advanced Artificial Intelligence, Larry M. Arjomandi and Nemai Chandra Karmakar

Control Components Using Si, GaAs, and GaN Technologies, Inder J. Bahl

Design of Linear RF Outphasing Power Amplifiers, Xuejun Zhang, Lawrence E. Larson, and Peter M. Asbeck

Design Methodology for RF CMOS Phase Locked Loops, Carlos Quemada, Guillermo Bistué, and Iñigo Adin

Design of CMOS Operational Amplifiers, Rasoul Dehghani

Design of RF and Microwave Amplifiers and Oscillators, *Second Edition*, Pieter L. D. Abrie

Digital Filter Design Solutions, Jolyon M. De Freitas

Discrete Oscillator Design Linear, Nonlinear, Transient, and Noise Domains, Randall W. Rhea

Distortion in RF Power Amplifiers, Joel Vuolevi and Timo Rahkonen

Distributed Power Amplifiers for RF and Microwave Communications, Narendra Kumar and Andrei Grebennikov

Electric Circuits: A Primer, J. C. Olivier

Electronics for Microwave Backhaul, Vittorio Camarchia, Roberto Quaglia, and Marco Pirola, editors

An Engineer's Guide to Automated Testing of High-Speed Interfaces, Second Edition, José Moreira and Hubert Werkmann

Envelope Tracking Power Amplifiers for Wireless Communications, Zhancang Wang

Essentials of RF and Microwave Grounding, Eric Holzman

Frequency Measurement Technology, Ignacio Llamas-Garro, Marcos Tavares de Melo, and Jung-Mu Kim

FAST: Fast Amplifier Synthesis Tool—Software and User's Guide, Dale D. Henkes

Feedforward Linear Power Amplifiers, Nick Pothecary

Filter Synthesis Using Genesys S/Filter, Randall W. Rhea

Foundations of Oscillator Circuit Design, Guillermo Gonzalez

Frequency Synthesizers: Concept to Product, Alexander Chenakin

Fundamentals of Nonlinear Behavioral Modeling for RF and Microwave Design, John Wood and David E. Root, editors

Generalized Filter Design by Computer Optimization, Djuradj Budimir

Handbook of Dielectric and Thermal Properties of Materials at Microwave Frequencies, Vyacheslav V. Komarov

Handbook of RF, Microwave, and Millimeter-Wave Components, Leonid A. Belov, Sergey M. Smolskiy, and Victor N. Kochemasov

High-Efficiency Load Modulation Power Amplifiers for Wireless Communications, Zhancang Wang

High-Linearity RF Amplifier Design, Peter B. Kenington

High-Speed Circuit Board Signal Integrity, Second Edition, Stephen C. Thierauf

Integrated Microwave Front-Ends with Avionics Applications, Leo G. Maloratsky

Intermodulation Distortion in Microwave and Wireless Circuits, José Carlos Pedro and Nuno Borges Carvalho

Introduction to Modeling HBTs, Matthias Rudolph

An Introduction to Packet Microwave Systems and Technologies, Paolo Volpato

Introduction to RF Design Using EM Simulators, Hiroaki Kogure, Yoshie Kogure, and James C. Rautio

Introduction to RF and Microwave Passive Components, Richard Wallace and Krister Andreasson

Klystrons, Traveling Wave Tubes, Magnetrons, Crossed-Field Amplifiers, and Gyrotrons, A. S. Gilmour, Jr.

Lumped Element Quadrature Hybrids, David Andrews

Lumped Elements for RF and Microwave Circuits, Second Edition, Inder J. Bahl

Microstrip Lines and Slotlines, Third Edition, Ramesh Garg, Inder Bahl, and Maurizio Bozzi

Microwave Component Mechanics, Harri Eskelinen and Pekka Eskelinen

Microwave Differential Circuit Design Using Mixed-Mode S-Parameters, William R. Eisenstadt, Robert Stengel, and Bruce M. Thompson

Microwave Engineers' Handbook, Two Volumes, Theodore Saad, editor

Microwave Filters, Impedance-Matching Networks, and Coupling Structures, George L. Matthaei, Leo Young, and E. M. T. Jones

Microwave Imaging Methods and Applications, Matteo Pastorino and Andrea Randazzo

Microwave Material Applications: Device Miniaturization and Integration, David B. Cruickshank

Microwave Materials and Fabrication Techniques, Second Edition, Thomas S. Laverghetta

Microwave Materials for Wireless Applications, David B. Cruickshank

Microwave Mixer Technology and Applications, Bert Henderson and Edmar Camargo

Microwave Mixers, Second Edition, Stephen A. Maas

Microwave Network Design Using the Scattering Matrix, Janusz A. Dobrowolski

Microwave Power Amplifier Design with MMIC Modules, Howard Hausman

Microwave Radio Transmission Design Guide, Second Edition, Trevor Manning

Microwave and RF Semiconductor Control Device Modeling, Robert H. Caverly

Microwave Transmission Line Circuits, William T. Joines, W. Devereux Palmer, and Jennifer T. Bernhard

Microwave Techniques in Superconducting Quantum Computers, Alan Salari

Microwaves and Wireless Simplified, Third Edition, Thomas S. Laverghetta

Millimeter-Wave GaN Power Amplifier Design, Edmar Camargo

Modern Microwave Circuits, Noyan Kinayman and M. I. Aksun

Modern Microwave Measurements and Techniques, Second Edition, Thomas S. Laverghetta

Modern RF and Microwave Filter Design, Protap Pramanick and Prakash Bhartia

Neural Networks for RF and Microwave Design, Q. J. Zhang and K. C. Gupta

Noise in Linear and Nonlinear Circuits, Stephen A. Maas

Nonlinear Design: FETs and HEMTs, Peter H. Ladbrooke

Nonlinear Microwave and RF Circuits, Second Edition, Stephen A. Maas

On-Wafer Microwave Measurements and De-Embedding, Errikos Lourandakis

Parameter Extraction and Complex Nonlinear Transistor Models, Günter Kompa

Passive RF Component Technology: Materials, Techniques, and Applications, Guoan Wang and Bo Pan, editors

PCB Design Guide to Via and Trace Currents and Temperatures, Douglas Brooks with Johannes Adam

Practical Analog and Digital Filter Design, Les Thede

Practical Microstrip Design and Applications, Günter Kompa

Practical Microwave Circuits, Stephen Maas

Practical RF Circuit Design for Modern Wireless Systems, Volume I: Passive Circuits and Systems, Les Besser and Rowan Gilmore

Practical RF Circuit Design for Modern Wireless Systems, Volume II: Active Circuits and Systems, Rowan Gilmore and Les Besser

Principles of RF and Microwave Design, Matthew A. Morgan

Production Testing of RF and System-on-a-Chip Devices for Wireless Communications, Keith B. Schaub and Joe Kelly

Q Factor Measurements Using MATLAB, Darko Kajfez

Radio Frequency Integrated Circuit Design, Second Edition, John W. M. Rogers and Calvin Plett

Relativistic Field Theory for Microwave Engineers, Matthew A. Morgan

Reflectionless Filters, Matthew A. Morgan

RF Bulk Acoustic Wave Filters for Communications, Ken-ya Hashimoto

RF Circuits and Applications for Practicing Engineers, Mouqun Dong

RF Design Guide: Systems, Circuits, and Equations, Peter Vizmuller

RF Linear Accelerators for Medical and Industrial Applications, Samy Hanna

RF Measurements of Die and Packages, Scott A. Wartenberg

The RF and Microwave Circuit Design Handbook, Stephen A. Maas

RF and Microwave Coupled-Line Circuits, Rajesh Mongia, Inder Bahl, and Prakash Bhartia

RF and Microwave Oscillator Design, Michal Odyniec, editor

RF Power Amplifiers for Wireless Communications, Second Edition, Steve C. Cripps

RF Systems, Components, and Circuits Handbook, Ferril A. Losee

Scattering Parameters in RF and Microwave Circuit Analysis and Design, Janusz A. Dobrowolski

The Six-Port Technique with Microwave and Wireless Applications, Fadhel M. Ghannouchi and Abbas Mohammadi

Solid-State Microwave High-Power Amplifiers, Franco Sechi and Marina Bujatti

Stability Analysis of Nonlinear Microwave Circuits, Almudena Suárez and Raymond Quéré

Substrate Integrated Suspended Line Circuits and Systems, Kaixue Ma and Yongqiang Wang

Substrate Noise Coupling in Analog/RF Circuits, Stephane Bronckers, Geert Van der Plas, Gerd Vandersteen, and Yves Rolain

System-in-Package RF Design and Applications, Michael P. Gaynor

Technologies for RF Systems, Terry Edwards

Terahertz Metrology, Mira Naftaly, editor

Understanding Quartz Crystals and Oscillators, Ramón M. Cerda

Vertical GaN and SiC Power Devices, Kazuhiro Mochizuki

The VNA Applications Handbook, Gregory Bonaguide and Neil Jarvis

Wideband Microwave Materials Characterization, John W. Schultz

*Wired and Wireless Seamless Access Systems for Public Infrastructur*e, Tetsuya Kawanishi

For further information on these and other Artech House titles, including previously considered out-of-print books now available through our In-Print-Forever® (IPF®) program, contact:

Artech House
685 Canton Street
Norwood, MA 02062
Phone: 781-769-9750
Fax: 781-769-6334
e-mail: artech@artechhouse.com

Artech House
16 Sussex Street
London SW1V 4RW UK
Phone: +44 (0)20 7596 8750
Fax: +44 (0)20 7630 0166
e-mail: artech-uk@artechhouse.com

Find us on the World Wide Web at: www.artechhouse.com