Restructuring the introductory electricity and magnetism course

Ruth Chabay[a)] and Bruce Sherwood
Department of Physics, North Carolina State University, Box 8202, Raleigh, North Carolina 27695

(Received 20 September 2005; accepted 12 December 2005)

In the electricity and magnetism (E&M) segment of the traditional introductory calculus-based physics course, many new and increasingly abstract concepts, embodied in complex formal relations, are introduced at a rapid pace. As a result, many students find E&M significantly more difficult than classical mechanics. We describe a different intellectual structure for the E&M course that stresses conceptual coherence, connects the abstract field concept to concrete microscopic models of matter, and follows a clear story line, culminating in the classical model of the interaction of electromagnetic radiation and matter. This sequence has proven to be effective in teaching the basic concepts of E&M.
[DOI: 10.1119/1.2165249]

I. WHY IS ELECTRICITY AND MAGNETISM DIFFICULT FOR STUDENTS?

Because electromagnetic interactions play a central role in determining the structure of the natural world and are the foundation of most current and emergent technology, a basic understanding of electricity and magnetism (E&M) is important. Traditionally, science and engineering students are introduced to E&M in the second half of the introductory calculus-based physics course, after they have completed an introduction to classical mechanics. However, even students who have done well in the first part of the course often find E&M to be difficult and confusing.

In E&M students encounter for the first time a level of abstraction and mathematical sophistication far beyond what they have experienced. In mechanics many situations involve familiar macroscopic objects: balls and sticks, cars and airplanes. At least some important concepts, such as velocity and force, are easily related to everyday experience. In E&M the student is quickly introduced to a world in which almost all of the quantities are invisible; they are either microscopic such as electrons or abstractions such as field, flux, and potential. Integral calculus becomes a central mathematical tool, and students are asked to apply it in unfamiliar ways, such as calculating the path integral or surface integral of a quantity expressed as a vector dot product. For the first time, it is necessary for students to think and visualize in three dimensions, a skill they have never before practiced. The role of symmetry and the nature of a symmetry argument as used in E&M are alien to students who have invested hundreds of hours in algebraic reasoning, but who have little experience with topology or with formal logical reasoning.

In the traditional introductory E&M sequence, the usual approach to this onslaught of new concepts and ways of reasoning is to gloss over it, going through the fundamentals at high speed, and spending most of the course on rote problem solving. The conceptual and mathematical complexity of the field is exacerbated by the extraordinarily rapid introduction of a long sequence of new and increasingly abstract concepts. The ideas of charge, electric force, field, flux, and Gauss's law are often presented within the first couple of weeks of the course. These ideas are quickly followed by the concepts of potential, potential difference, and electric current, which appear to be only slightly related to the previous set of concepts. Students can easily be overwhelmed by this rapid introduction of abstract ideas and usually are not given sufficient practice to be able to apply these concepts reliably, or to discriminate them from each other. By the end of the course, even good students may have forgotten the expression for the electric field of a single point charge, because it has not been used for many weeks. Students who can reliably solve complex circuit problems often believe at the end of the course that electrons are used up in light bulbs or that the current produced by a battery is independent of the circuit it drives.[1,2] The rapid introduction of new concepts and escalation in complexity frequently confirms in students' minds the conviction that physics consists of a large number of disconnected formulas.

II. GOALS OF THE INTRODUCTORY PHYSICS COURSE

Some research and development in physics education has focused on remedying particular problems with the traditional sequence by giving students additional focused practice on selected concepts. However, without addressing the overarching issues of structure and coherence, it is difficult to do more than improve student performance on isolated tasks. We have chosen instead to reexamine the intellectual structure of the E&M curriculum to identify which concepts are centrally important, how these concepts are related, and how they can be introduced to students at the introductory level in a coherent, comprehensible sequence. To structure a sequence whose intellectual coherence is evident to students as well as instructors, it is necessary to have a clear set of goals to guide decisions about which topics to include and when and how to present them.

We believe that an appropriate goal for the introductory calculus-based physics curriculum is to engage students in the contemporary physics enterprise by emphasizing the following:

- A small number of fundamental principles.
- A unified approach based on the integration of the atomic nature of matter and macro/micro connections.
- The modeling of real physical systems, including computational modeling.

Students who have completed the introductory calculus-based physics course should understand that a small number of fundamental principles can explain a very wide range of phenomena—a central goal of the curriculum. Students should be capable of applying fundamental principles to new problems. They should clearly understand the power of classical and semiclassical models even at a microscopic scale and be aware of the limitations of purely classical, macroscopic models. Students should be able to decide when it is appropriate to introduce simplifying assumptions and approximations and be able to use approximations productively, including the use of numerical methods.[3–5]

In contrast, the traditional rationale for introductory physics is to develop systematic problem-solving skills and to give students practice in applying mathematics. Little effective attention is given to the broader goal of bringing students to appreciate the unity of physics and the power of a limited number of fundamental principles. Although textbook authors and physics instructors may have this reductionist nature of physics in mind, the standard organization and content of the course and the way that it usually progresses confirms in many students the conviction that their task is to remember which formula is to be used in a particular problem.

There are several ways to restructure the introductory E&M course to provide greater conceptual coherence. In this paper we describe one such restructuring and discuss several aspects of the sequence and content that contribute to greater conceptual coherence and facilitate deeper learning of the basic concepts.

III. CONTENT, SEQUENCE, AND EMPHASIS

The goals of the new sequence are to increase conceptual coherence, give students time to assimilate and master new concepts, add concreteness, and help students to develop microscopic models that facilitate reasoning about complex systems. The organization of topics is hierarchical, and the overarching theme of the entire sequence is the field concept. The sequence is organized into four large segments:

- Stationary charges
  - Electric field
  - A microscopic model of matter (conductors and insulators)
  - Effect of electric field on matter; the approach to equilibrium
  - Electric field of distributed charges
  - Electric energy and electric potential
- Moving charges
  - Magnetic field
  - Microscopic view of circuits (charge, field, energy, and the potential in dc and RC circuits)
  - Macroscopic view of circuits
  - Magnetic force; microscopic view of magnetic forces on currents
- Reasoning about patterns of field in space
  - Gauss's law and Ampere's law
- Time-varying fields and accelerated charges
  - Faraday's law
  - Maxwell's equations; electromagnetic radiation; classical interaction of light and matter
  - Physical optics; wave-particle duality

In the following, we discuss particular aspects of this sequence and indicate how these topics fit together into a coherent structure.

IV. FIELD: AN INTERMEDIATE LEVEL OF ABSTRACTION

The concept of field is central to electricity and magnetism. In the traditional introductory course this concept is not used during large sections of the course, including the sections dealing with electric circuits and Faraday's law. Consequently, the field concept does not appear central to students, who are kept busy plugging numbers into formulas for equivalent resistance and mutual inductance. A goal of the redesigned topic sequence is to make the field concept appear more important, comprehensible, and useful to the students.

The field concept is a significantly more abstract concept than the quantities typically encountered in mechanics. However, the effort required to understand fields is justified by the immense gain in predictive and explanatory power it affords. In particular, the classical model of electromagnetic radiation is incomprehensible to a student who has not mastered the concepts of electric and magnetic fields. Like the Newtonian synthesis in mechanics, Maxwell's equations and the classical explanation of the nature of light are one of the crowning intellectual achievements of classical physics. Introductory students are capable of understanding this triumph if and only if they have had sufficient practice and sufficiently varied experiences with electric and magnetic fields and the effects of these fields on matter, before they encounter the even more complex concept of an electromagnetic wave. For this reason, and because the field concept alone affords students the opportunity to gain significantly in intellectual sophistication, the field concept is an appropriate central focus of the introductory E&M sequence and is the backbone of the story line of the course.

One of the ways in which the field concept can be made concrete and connected to the behavior of matter is by an emphasis on the crucial role of dipoles, both electric and magnetic, permanent and induced. The fields made by dipoles and the creation of dipoles by applied fields play a significant role throughout the new sequence.

A. Magnetic field

In the traditional sequence magnetic field and magnetic force are introduced late in the course, after electrostatics and circuits. This delay has many disadvantages. The rapid introduction of both magnetic field and magnetic force makes both concepts difficult to assimilate, because they involve vector cross products and require difficult mental rotations. Students have little remaining time in the course to gain adequate experience with the topic and little time to compare and contrast electric and magnetic fields and their effects. When Faraday's law, which involves the time variation of magnetic flux, is introduced immediately afterward, the conceptual and mathematical complexity escalates significantly.

Experience with magnetic field and magnetic force is necessary to solidify students' understanding of the general concept of a field. When we introduced magnetic field and magnetic force in the traditional sequence, we observed that even strong students were frequently disconcerted to realize that they had previously missed the central two-step nature of analyzing interactions mediated by fields (charges make fields, fields affect other charges). When dealing with electric interactions, it is possible for students to imagine distributions of source charges interacting directly with other charges and to reason using the Coulomb force law, instead of focusing on the field as a mediator of the interaction. However, the complex spatial nature of magnetic interactions leaves no alternative to the two-step process of determining the magnetic field due to the source charges, then using this field to find the direction and magnitude of magnetic forces.

Besides strengthening students' understanding of fields, the early introduction of the magnetic field allows students to use the magnetic field of moving charges as a probe of the current in electric circuits. Initially, the magnetic field can be viewed operationally as a field that affects a compass. A magnetic field is produced by moving charges, and students can observe the magnetic effects that currents in simple circuits have on a nearby compass. They can observe the magnetic field near a coil and a bar magnet and identify a dipole-like pattern familiar from earlier work with electric dipoles. An atomic model allows students to predict the magnetic dipole moment of a bar magnet and to confirm their prediction by using a compass to measure it. The treatment of magnetic force can also make strong connections to the electric field and force and to post-classical physics. For example, a discussion of two moving protons, viewed from two different reference frames, can clearly show how electric and magnetic effects are dependent on the choice of reference frame and can offer the opportunity to explore the idea that time runs differently in different reference frames in a context accessible to students at this level.

As discussed in a later section, both electric and magnetic fields can play a central role in the analysis of electric circuits. As a result of this continued exposure, by the time Faraday's law and electromagnetic radiation are introduced, students have had many weeks of experience with both electric and magnetic fields and in the analysis of situations in which both fields play important roles.

B. Effects of fields on matter

The restructured sequence emphasizes the atomic nature of matter. Fields by themselves are very abstract, but our experience with fields comes from observing their effect on material objects. In the traditional curriculum polarization phenomena are often briefly mentioned, presumably because atoms and their constituent particles are not discussed, so reasoning about these complex phenomena is difficult. In the new approach the polarization of salt solutions, metals, molecules, and insulating solids is discussed in detail. To do so, we need simple microscopic models of atoms and molecules and especially of solids. Cognitive scientists have found that it is significantly easier for people to reason by running mechanistic mental models of processes than by engaging in the kind of formal, global, constraint-based reasoning common in macroscopic classical physics.[6] To make sense of the response of material objects to electric and magnetic fields, it is important to have a simple model of the atomic nature of matter, especially solids. For example, the attraction of a neutral object, whether a conductor or an insulator, to an object with a nonzero net charge of either sign can be understood only by thinking about the effect of applied fields on the electrons in the neutral material.

Students in the introductory physics course have usually learned about liquids and gases in previous chemistry courses, but the structure of solids, especially metals, is typically not covered in introductory physics or chemistry courses. We have observed that even students who are familiar with the ball and spring model of a solid (from a preceding mechanics course) are typically surprised by the existence of a mobile electron sea in metals. The discussion of this model lays an important foundation for later work, including the microscopic analysis of circuit behavior discussed in Sec. V and the classical model of the interaction of light and matter. In addition, a microscopic model supports the discussion of the transients involved in polarization, making it possible to reason step-by-step about the processes involved in the approach to static equilibrium (or later, in circuits, the approach to the steady state, or the quasi-steady state in RC circuits). This focus on transient processes leads to discussions of the role of retardation, setting the stage for later simple explorations of fields in moving reference frames. Retardation and other relativistic effects are important in many aspects of modern science and technology and introductory E&M can provide a relatively easy introduction to these concepts.

V. FIELD, MICROSCOPIC MODELS OF MATTER, AND ELECTRIC CIRCUITS

The analysis of electric circuits provides an opportunity to solidify the new concepts that have been introduced (charge, electric field, potential, magnetic field, models of matter). However, in the traditional E&M curriculum electrostatics and circuits are treated as almost completely separate topics. Although electrostatic phenomena are analyzed in terms of charge and field, circuits are analyzed in terms of current and potential, and the connection between these two sets of concepts is not made salient. This dissociation can reinforce the perception that physics consists of a large number of special case formulas. In addition, this approach removes the concept of electric field from the student's view, so that by the end of the course students might have forgotten most of what they learned about the concept of electric field in the beginning of the course. To stress the fundamental nature of the field concept and the microscopic view of matter, both dc and RC circuits can be analyzed from a microscopic point of view in terms of the electric field and the microscopic properties of conductors.

Many pathologies in reasoning can be avoided by dealing in approximations rather than infinities. For example, instead of resistanceless wires, one can discuss wires with small but finite resistance. It then makes sense that a nonzero electric field must be established inside such a wire to keep charged particles moving and to sustain a current. The key to understanding the source of the electric field at the microscopic level is the surface-charge model of circuits. This model has been discussed for many years but has rarely been mentioned in introductory textbooks.[7] Haertel's monograph[8] points out the explanatory power of this model.

From a microscopic point of view, ohmic materials are described in terms of the microscopic relation $\nu = \mu E$, where ν is the drift speed, μ is the mobility of the mobile charges, and E is the electric field inside the material. The field inside circuit wires can be shown to be due not only to charges in and on the battery but also to those on the surfaces of the circuit elements. Students learn to analyze both dc and RC circuits directly in terms of the Coulomb interaction and the atomic nature of matter. This analysis in terms of charge and field makes a strong connection with electrostatics, unifies the two topics, and provides a strong sense of mechanism for the behavior of simple circuits.

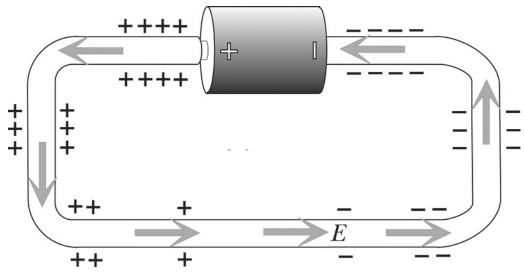


Fig. 1. The steady state pattern of the electric field and a schematic surface charge distribution for a simple circuit in the steady state. The surface charge is shown at sample locations along the wire to depict a charge gradient.

Figure 1 shows the pattern of the electric field and a schematic charge distribution on a simple circuit consisting of a battery connected to a resistive wire of constant cross section and composition. In the steady state there is a gradient of surface charge density on the surface of the resistive wire, which is a major contributor to the field inside the wire (charges in and on the battery also contribute). In a long straight current-carrying coaxial cable the surface charge density on the central conductor has a constant gradient,[9,10] but in the geometry of Fig. 1 a constant gradient is only an approximation to the actual charge distribution. What we do know precisely is that in the steady state the charge distribution that is established is the one that produces the known simple pattern of electric field in the wire, which is uniform in magnitude and everywhere parallel to the wire.

In an ordinary 3 V circuit the surface charge is too small to produce observable electrostatic effects; the fields in the circuit elements are small compared with the fields observed in electrostatic phenomena. The field in a 1 m resistive wire is only 3 V/m compared to the field of 3×10^6 V/m in air that initiates a spark. However, in a 10 kV circuit the surface charge is large enough to have observable mechanical consequences. In Fig. 2 we show a demonstration suggested in Ref. 11. The thick wires represent bare copper conducting wires, and the thin wires represent high-resistance resistors. At position A there is a large negative surface charge density, at B there is less, and at C the surface charge density is zero by symmetry. This distribution of surface charge produces large fields in the resistors. If we bring a piece of aluminized mylar near position A, it is observed to be strongly attracted to the bare wire and then strongly repelled after charging by contact. A test for sign shows the mylar to be charged negatively. At B the effect is less strong and at C nothing happens. Similar behavior is noted at positions D and E, but the

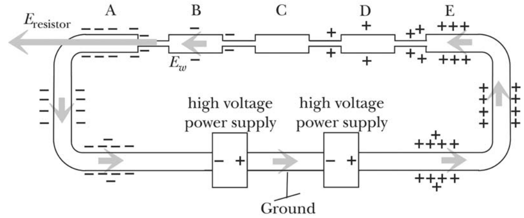


Fig. 2. A high voltage, low current circuit in which the buildup of surface charge is large enough to be easily detectable. The thick sections represent thick connecting wires; the four thin segments are high-resistance resistors. The surface charge gradient is shown schematically.

sign is found to be positive. This demonstration is a dramatic illustration of the reality of the surface charge that contributes to the field in the conductors that drives the current in the circuit.

Through the surface charge model students can acquire a deep sense of the mechanism for circuit behavior, including the transient in which the steady state is established through feedback. The subsequent connection of this model to the traditional macroscopic analysis of circuits allows students to reinforce the connections between field and potential difference and to see the microscopic components of macroscopic quantities such as resistance. For further discussion, including a historical overview, see Ref. 12.

VI. PATTERNS OF FIELD IN SPACE: GAUSS's LAW

In the traditional sequence Gauss's law is introduced very early, sometimes during the first week of the course. Generations of physics teachers have lamented the fact that the students do not understand Gauss's law. From a cognitive point of view it is clear why this lack occurs despite the best efforts of good teachers. At the beginning of the course, many students are struggling with what is for them a subtle distinction between charge and field, yet Gauss's law embodies a complex topological relationship between charge and patterns of field in three-dimensional space. Early in the course students have had no experience with the kinds of patterns of field that are possible in space, but these patterns of field lie at the heart of the topological relationship. Surface integrals are typically unfamiliar to students, who have usually not completed a multivariable calculus course before starting E&M, so flux (which involves a vector dot product inside a surface integral) is a challenging concept. Moreover, students have little experience visualizing three-dimensional geometries and little experience with symmetry arguments of the kind that play a major role in the applications made of Gauss's law. The shakiness of the concepts of charge and field, the students' lack of experience with possible field configurations, their lack of mathematical background, and their unfamiliarity with symmetry arguments make the introduction of Gauss's law early in the course a frustrating endeavor.

Why is Gauss's law traditionally introduced early? The justification is that it is needed to prove two important properties of electrostatics, that excess charge is found only on the surface of a conductor (not in the interior), and that the electric field inside an empty cavity in a conductor in static equilibrium is zero everywhere. Probably the reason Gauss's law has traditionally been treated early is to satisfy the desire for rigor on the part of the teacher. Unfortunately, at this point it is only the instructor who is able to appreciate this rigor. Because it poses such conceptual challenges for students, it is appropriate to delay the introduction of Gauss's law until much later in the course. How then can these results be introduced when they are needed, near the beginning of the course? We can make both results plausible and state that we will later be able to derive them rigorously when Gauss's law is discussed. The mutual repulsion of electrons suggests but does not prove that excess charges will appear on the surface of a conductor (this is true only because the electric force is an inverse square force). If the interior of a conductor is neutral, the removal of a neutral chunk of the interior should not change the net field inside, and it has

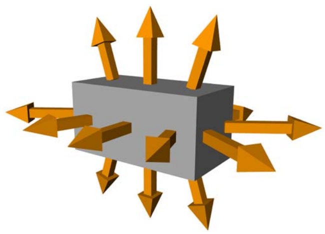


Fig. 3. Electric field (arrows) at locations on the surface of an opaque Gaussian "box."

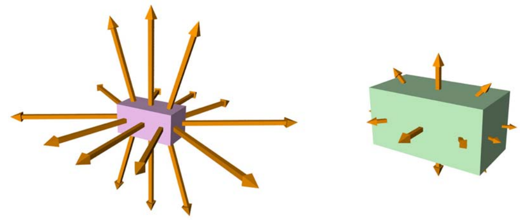


Fig. 5. The electric field (arrows) shown at locations on a small Gaussian surface (left) and a larger surface (right), both enclosing the same charge distribution.

already been proven that the net electric field inside a conductor in static equilibrium is zero. Students are quite willing to accept these plausible arguments.

Gauss's law is also typically used early in the course to derive the electric fields of a uniformly charged spherical shell, an infinite uniformly charged plate, and an infinite uniformly charged rod. However, it is not actually necessary to invoke Gauss's law to obtain these results. The electric field of a uniformly charged spherical shell can be derived directly using calculus; one can show how to set up the integral over the charge distribution but explain that an easier route to the final result will be available later using Gauss's law. The infinite uniformly charged plate and an infinite uniformly charged rod are limiting cases of the finite disk and finite rod, which are derived fairly easily by integration. There is a pedagogical advantage to emphasizing large but finite charge distributions rather than infinite ones, because the latter can raise awkward conceptual issues. (For example, a student correctly pointed out to us that it would take infinitely long to charge a capacitor with infinite plates.)

In the new sequence Gauss's law is introduced about two-thirds of the way through the course when students have had much experience with patterns of electric (and magnetic) fields in different contexts, including in electric circuits. This experience can now be exploited to introduce Gauss's law first in a qualitative, visual form that emphasizes what students already know about patterns of fields in space and the charge distributions responsible for them. For example, if students are shown a diagram such as the one in Fig. 3, which depicts measurements of electric field (arrows) at locations on the surface of an imaginary opaque box-shaped surface, they have little trouble at this point in the course concluding that there must be positive charge inside the surface.

With a bit more thought, students are also able to conclude that there is positive charge inside the second closed surface shown in Fig. 4. Students typically reason about this second situation by envisioning charge distributions that would produce this pattern of electric field in a region, eventually concluding that a large uniform positively charged plate passing through the center of the box could be responsible.

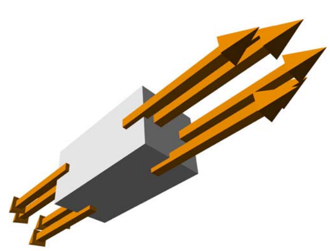


Fig. 4. Electric field (arrows) at locations on a closed Gaussian surface (gray box).

In lectures, the use of dynamic, navigable three-dimensional (3D) computer visualization software[12] that aids in visualizing Gaussian surfaces and inspecting the field pattern on the surfaces can be particularly helpful to the large fraction of students who still find imagining 3D patterns challenging.

When students begin with a qualitative understanding of the topological connection between patterns of fields in space and the distribution of source charges, the formal proof of Gauss's law becomes significantly more accessible to them. Key pieces of the proof can also be introduced qualitatively; for example, the proportionality of the magnitude of the measured field to the amount of charge inside the closed surface, the relation of the outwardly normal component of field to the sign of the enclosed charge, and the necessity of taking into account the surface area of the imaginary closed surface (as in Fig. 5, in which the enclosed charge is the same, but the size of the Gaussian surface, and hence the magnitude of the field at locations on the surface is different).

If the qualitative, topological nature of Gauss's law is clear to students, they are better able to apply Gauss's law to new situations as well as to understand the usual proofs and applications (for example, no excess charge in the interior of a conductor, zero net electric field inside a hollow metal object). Students who have analyzed circuits from a microscopic viewpoint can subsequently apply Gauss's law to find the amount and sign of the charge on the interface between two materials with different conductivities in a steady-state circuit, as shown in Fig. 6.

Ampere's law, which, like Gauss's law, involves patterns of field in space, is treated in the same fashion, building on students' familiarity with the patterns of magnetic field produced by typical configurations of currents and bar magnets.

VII. FARADAY'S LAW

Faraday's law is usually difficult for students. It involves a dynamic connection between magnetic and electric phenomena and is traditionally introduced when students have had only a rather brief exposure to magnetic fields and when the electric field concept has not recently been used. Moreover, the integral form (which is the usual form introduced in the introductory course, because most students have not yet encountered divergence and curl in their calculus courses) involves the concept of flux, which is traditionally introduced at the start of the course in the context of Gauss's law and not mentioned again until the introduction of Faraday's law.

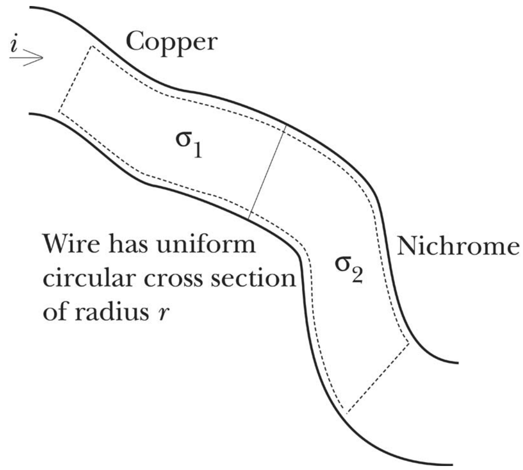


Fig. 6. Use of Gauss's law to find the amount and sign of excess charge buildup on the interface between two cylindrical conductors with the same dimensions but different conductivities. The dotted line indicates a Gaussian surface with circular cross section.

The effect is to use a forgotten concept (flux) to relate a line integral of electric field (emf) to the time derivative of a surface integral of a quantity with which the students have had inadequate practice (magnetic field). It is not surprising that Faraday's law is usually difficult for students.

Faraday's law can be introduced with an emphasis on the curly electric field that is found surrounding a region of time-varying magnetic field. Students at this point have had extensive experience with both electric and magnetic fields, including many examples of patterns of these fields in space. The flux concept is fresh in their minds, because the discussion of Gauss's law immediately precedes Faraday's law. The combination of long-term experience with electric and magnetic fields with the just-in-time treatment of the flux concept can make Faraday's law much more accessible to students.

Although it is relatively easy to introduce magnetism early and refer to it repeatedly, it would be difficult to find appropriate applications of the flux concept to sustain and build student facility with the concept, even if it were possible to teach Gauss's law effectively at the start of the course. Hence there is another advantage to the just-in-time introduction of the concept of electric flux, immediately preceding Faraday's law. It is worth noting that motional emf can usefully be discussed much earlier in the context of magnetic force, well before the introduction of Faraday's law. Introducing motional emf before encountering Faraday's law helps students make an important distinction between these two very different mechanisms for producing emf (magnetic force on moving charged particles versus a time varying magnetic field), which often are not clearly differentiated in the traditional sequence.

VIII. ELECTROMAGNETIC RADIATION

After discussing Gauss's law for electricity and magnetism, Ampere's law, and Faraday's law, one is ready to consider Maxwell's extension to Ampere's law and show that crossed electric and magnetic fields can propagate in empty space at the speed of light. By using a qualitative version of an argument due to Purcell,[13] animated diagrams are used to show the results of retardation and make it plausible that an accelerated charge produces transverse radiative fields. The equation for the radiative fields of an accelerated charge may be stated without proof at this level. A sense of the mechanism for the production of radiation is important in making accessible the classical interaction of electromagnetic fields with matter, especially re-radiation. With a continual emphasis on the effects of fields on charged particles, it is natural to talk about the acceleration of the electrons in matter by the electric field in incident radiation and the subsequent re-radiation by these accelerated electrons. This view can bring to physical optics a clear sense of mechanism.

IX. MINIMALISM AND CHOICE OF REPRESENTATION: FIELD LINES

Because of the many new concepts in E&M, it is important to take a minimalist approach and to consider carefully the cost of introducing still more concepts and representations. In this spirit we advocate eliminating field lines from the course. At the introductory level there are almost no problems in which students can use field lines to reason about some phenomenon, and a significant investment of instructional time and a significant amount of practice are required if we are to expect students to interpret field lines correctly. Students are rarely taught to construct field line diagrams quantitatively, so this representation never becomes a really useful tool for them.

Many introductory textbooks introduce field lines, but even in these books the only homework problems requiring field lines are self-referential: Can they cross? Are they the trajectories of charged particles? Is there a field between the lines? Sometimes students are taught to count field lines to quantify the flux in Gauss's law problems; but because they are not able to draw correct field line diagrams, this approach is of limited value. The absence of authentic tasks is a sufficient reason to refrain from making an investment in teaching this topic; 3D vector diagrams can adequately depict patterns of fields in 3D space. There is not a serious conceptual cost, because at the introductory level field vectors provide all the necessary analysis tools. Because vectors are used continually throughout the course, a large investment in learning to use vectors to represent fields is necessary in any case.

Because there are several important misconceptions that appear to be suggested by the field line representation, the introduction of field lines requires a significant investment in confronting and attempting to dispel these ideas. These conceptual confusions include the belief that charged particles travel along field lines and that the lines are concrete objects that can interfere with the motion of other objects.[14] We have seen students draw field lines around a bar magnet in a pattern of tight curves running along the magnet from one end to the other and conclude from their drawing that there is no field on the axis of the magnet, where the field is in fact the largest. The field line representation can give confusing information about dynamic range to novices. For example, textbook field-line diagrams of the Earth's magnetic field sometimes give little sense of how very fast the field falls off with distance.

The typical field line diagram in two dimensions has scientific flaws, and in some quite ordinary cases it is impossible to draw such a diagram correctly.[15] For example, consider a circle of point charges in a plane and try to draw field lines inward from these charges. In two dimensions the lines must stop and cannot terminate on another charge. A correct representation requires flux tubes in three dimensions, not lines in two dimensions. The only serious way to deal with these problems is to invest much time in teaching about the limitations of two-dimensional field lines and combating the associated misconceptions. However, in the absence of authentic introductory-level tasks that require the use of this concept, the introduction of field lines is not a cost-effective investment of time and effort.

At a more advanced level, field lines are used extensively in reasoning about the complex phenomena that occur when fields and plasmas are strongly coupled. These applications are beyond the scope of the introductory E&M course, and discussions of clashing field lines are unlikely to help beginning students understand the basic aspects of electromagnetic fields.

X. THE TRANSITION FROM MECHANICS TO E&M

Changes in the content and emphasis of the mechanics course can facilitate the transition from mechanics to E&M. In our approach to mechanics, students are given practice in starting from a small number of fundamental principles and in working with simple microscopic models of matter.[3–5] Macro/micro connections in mechanics help students understand both macroscopic phenomena and microscopic phenomena. Along with gravitational forces and gravitational potential energy, electric force and electric potential energy are introduced and used routinely in the mechanics course. For example, a mechanics homework problem asks students to determine the required initial kinetic energy for a proton and deuteron to come into contact so they can fuse to give He^3 plus a photon; students also calculate the net energy gain in the fusion reaction. As a result, students beginning the E&M course have already had useful experience with electric interactions and microscopic models of matter.

XI. ASSESSMENT OF STUDENT LEARNING

Various studies have compared the performance of students who have completed the revised E&M sequence with the performance of students who have taken a traditional introductory E&M course. Of necessity, these comparisons have been restricted to standard topics covered in both courses and do not measure how well students have learned the material that is unique to the new sequence.

Thacker, Ganiel, and Boys found that students in the revised E&M sequence were able to solve difficult problems involving RC circuits significantly better than did students in a traditional curriculum.[16] Students in the revised curriculum were also able to give better explanations of their reasoning, which was often microscopic and mechanistic, while the other students relied on algebraic manipulation of formulas and were less frequently correct. Engelhardt and Beichner found that students from a course using the revised sequence outperformed students from a traditional course on a test of students understanding of dc circuits.[17]

We have conducted several studies comparing the performance of students who have studied the revised sequence in comparison to students who have taken traditional courses. One such study was a longitudinal study of what students learn and retain in the introductory E&M course and involved students who had completed the introductory E&M course at some time during the previous three years. The basic finding[18] was that on BEMA, a test of basic E&M concepts,[19] students in the revised version scored one letter grade better than did comparable students in a traditional course. In other words, students who received a B in a course using the revised sequence performed as well as students who received an A in a traditional course. There were no significant differences in aptitude or GPA between these groups. This effect persisted for five semesters after completing the course.

After verifying the reliability of BEMA as an assessment instrument,[20] it was used at North Carolina State University with engineering and science students in the second semester introductory calculus-based physics course, recruited from four traditional sections (with four different lecturers) and from four sections using the revised sequence (with four different lecturers). Students using the revised E&M sequence scored significantly higher than students in the traditional course, showing about twice the gain from the pretest level, which is a very large effect.[21] The two groups of students were otherwise indistinguishable (their distribution of GPAs and final grades in the physics and calculus courses were the same).

A different study involved complex problem solving.[22] Three traditional problems on a final exam in E&M were identical in two courses serving indistinguishable populations of students—one traditional, the other using the new sequence. The number of students who did each problem exactly correct was counted in each course. For two problems there was no significant difference. For the third problem, the most complex (consisting of many steps), the performance of the students who had studied the revised sequence was four times higher than that of the students in the traditional curriculum.

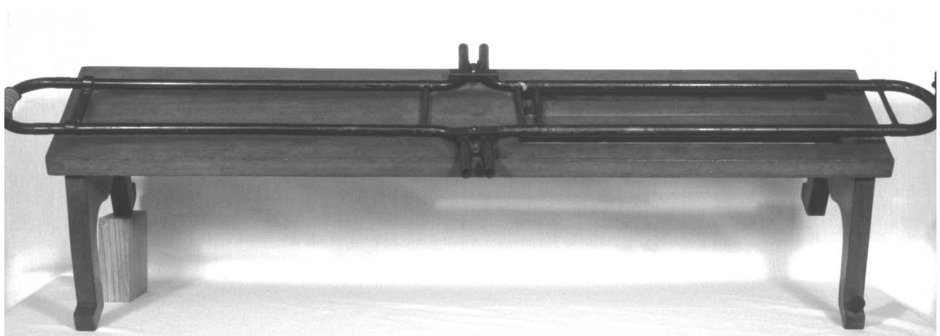
ACKNOWLEDGMENT

These developments were supported in part by NSF Grant No. DUE-0320608.

[a]Electronic mail: rwchabay@unity.ncsu.edu

[1]L. McDermott and P. Shaffer, "Research as a guide for curriculum development: An example from introductory electricity. I. Investigation of student understanding," Am. J. Phys. **60**(11), 994–1003 (1992); erratum **61**(1), 81 (1993).
[2]E. Mazur, *Peer Instruction* (Prentice Hall, Upper Saddle River, NJ, 1997), p. 5.
[3]R. Chabay and B. Sherwood, *Matter & Interactions I: Modern Mechanics and Matter & Interactions II: Electric & Magnetic Interactions* (Wiley, New York, 2002).
[4]R. Chabay and B. Sherwood, "Bringing atoms into first-year physics," Am. J. Phys. **67**(12), 1045–1050 (1999).
[5]R. Chabay and B. Sherwood, "Modern mechanics," Am. J. Phys. **72**(4), 439–445 (2004).
[6]P. N. Johnson-Laird, *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness* (Harvard U.P., Boston, 1986).
[7]N. Preyer, "Surface charges and fields of simple circuits," Am. J. Phys. **68**(11), 1002–1006 (2000). This paper contains extensive references to previous work.
[8]H. Haertel, *A Qualitative Approach to Electricity* (Institute for Research on Learning, Palo Alto, 1987).
[9]A. Sommerfeld, *Electrodynamics* (Academic, New York, 1952), pp. 125–130.
[10]A. Marcus, "The electric field associated with a steady current in long cylindrical conductor," Am. J. Phys. **9**(4), 225–226 (1941).
[11]U. Ganiel, private communication, January 1992.
[12]⟨http://www4.ncsu.edu/~rchabay/mi⟩.

[13]E. M. Purcell, *Electricity and Magnetism* (McGraw-Hill, New York, 1985), pp. 459–463.
[14]S. Törnkvist, K. A. Pettersson, and G. Tranströmer, "Confusion by representation: On student's comprehension of the electric field concept," Am. J. Phys. **61**(4), 335–338 (1993).
[15]A. Wolf, S. Van Hook, and E. Weeks, "Electric field line diagrams don't work," Am. J. Phys. **64**(6), 714–724 (1995).
[16]B. Thacker, U. Ganiel, and D. Boys, "Macroscopic phenomena and microscopic processes: Student understanding of transients in direct current electric circuits," Am. J. Phys. **67**(7), S25–S31 (1999).
[17]P. Engelhardt and R. Beichner, "Students' understanding of direct current resistive electrical circuits," Am. J. Phys. **72**(1), 98–115 (2004).
[18]R. Chabay and B. Sherwood, "Qualitative understanding and retention," AAPT Announcer **27**(2), 1050 (1997).
[19]BEMA (Brief E&M Assessment), available from the authors.
[20]L. Ding, R. Chabay, B. Sherwood, and R. Beichner, "Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment," Phys. Rev. ST Phys. Educ. Res. (to be published).
[21]L. Ding, R. Chabay, and B. Sherwood, "Applying BEMA (Brief Electricity & Magnetism Assessment) in calculus-based introductory E&M courses," AAPT Announcer **35**(2), 106 (2005).
[22]R. Chabay and B. Sherwood, "Qualitative understanding and problem-solving performance," AAPT Announcer **27**(2), 96 (1997).



Acoustic Interferometer. Acoustic interferometers operate by dividing the signal into two parts, which are sent on paths of different lengths and then recombined. If the path difference is an integer number of wavelengths, the waves interfere constructively when they rejoin, and a loud sound is heard. However, if the path length difference is a half-odd number of wavelengths, the signals are 180 degrees out of phase when they combine, and a minimum is produced. One side telescopes to vary the length of one of the paths. The sound is produced by a tuning fork of known frequency, and the maxima and minima are detected by using a rubber tube to lead the signal to the ear of the experimenter. This apparatus is at the University of Cincinnati. (Photograph and Notes by Thomas B. Greenslade, Jr., Kenyon College)