# A unified treatment of electrostatics and circuits

Bruce A. Sherwood
Center for Innovation in Learning and Dept. of Physics

Ruth W. Chabay
Center for Innovation in Learning

Carnegie Mellon University
Pittsburgh PA 15213

In introductory electricity courses, electrostatic and circuit phenomena are usually treated as separate and unrelated. By emphasizing the crucial role played by charges on the surfaces of circuit elements, it is possible to describe circuit behavior directly in terms of charge and electric field. This more fundamental description of circuits makes it possible to unify the treatment of electrostatics and circuits.

## I. INTRODUCTION

We have created a new course and textbook on electricity and magnetism for the introductory physics sequence, with an emphasis on qualitative reasoning based on dynamic atomic-level models.[1,2,3] An important component of the new course is a unified approach to electrostatics and circuits. This article explains the physics and pedagogy of this unified approach.

In the traditional syllabus for electricity, circuit phenomena are expressed solely in terms of the abstract concepts of potential and Kirchhoff's rules, making it seem that the fundamental Coulomb interaction plays little role in circuits. As a result, electrostatics and circuits appear to be two completely different and unrelated topics, and there is little learned in electrostatics that can be applied directly in analyzing circuits. Moreover, when electric field plays little or no role in circuit analysis, students lose an important opportunity for additional practice with electric field, and by the end of the discussion of circuits, students' understanding of the field concept may have faded.

Another problem with the traditional syllabus is that serious misconceptions often persist even after extensive instruction, such as the notion that current is used up in a light bulb. An approach to circuits that emphasizes a deeper sense of mechanism can address this problem.

The initial inspiration for unifying electrostatics and circuits came from the work of Haertel.[4,5] In this scheme, the basic Coulomb interaction expressed in terms of charge and electric field is sufficient to analyze both electrostatic and circuit phenomena in a unified way. After completing this unified treatment in terms of the Coulomb interaction, we introduce potential, and we then re-examine both electrostatics and circuits in terms of potential. Figure 1 compares the traditional and new treatments of circuits.

| Traditional treatment of circuits | New treatment of circuits |
|---|---|
| Little or no connection to electrostatics | Unified treatment of electrostatics and circuits |
| Solely in terms of potential and current | Initially in terms of charge and field (followed up later by analysis in terms of potential and current) |
| Macroscopic only | Microscopic as well as macroscopic |
| Steady state only | Transient polarization establishes the steady state |
| Little sense of mechanism | Strong sense of mechanism |

Figure 1. Comparison of traditional and new treatment of circuits.

The new curriculum deals in a natural way with many of the difficulties observed in learning electricity. The immediate and direct benefit is a unified treatment of electrostatics and electric circuits, which closes the observed gaps in students' reasoning about these topics. Furthermore, this treatment provides an improved basis for understanding circuits in terms of the more abstract concept of potential. In the traditional curriculum, students who can set up equations to solve for currents in complex multi-loop circuits nonetheless often have difficulty using potential to analyze even simple circuits qualitatively.[6,7] This is especially true in dynamic situations involving transients and when dealing with the consequences of introducing a change into an electric circuit (e.g., adding a resistor). The new model provides a concrete basis for reasoning dynamically about these concepts in the analysis of electric circuits.

Students' tendency to reason locally and sequentially about electric circuits[8,9] is directly addressed in this new approach. One analyzes dynamically the behavior of the *whole* circuit, and there is a concrete physical mechanism for how different parts of the circuit interact globally with each other, including the way in which a downstream resistor can affect conditions upstream.

Common student misconceptions directly addressed by the new approach include the following: current is used up in a light bulb; the electric field inside a metal is always zero (even when the system is not in static equilibrium); drifting electrons push each other through a wire just as water molecules push each other through a pipe (despite charge neutrality inside the metal); Ohm's law applies to all circuit elements (not just resistors); the Kirchhoff loop rule is identical to Ohm's law (instead of being a separate and much more general principle); there cannot be any potential difference across an open switch because V = IR, and there is no I; a battery is either on or off and constitutes a constant-current device; emf and potential difference are synonymous.

Initially it was not at all clear whether an integrated approach to circuits and electrostatics would be feasible in an introductory course. We experimented with various sequences of topics before finding a scheme that is successful. In absolute terms, we find that students are indeed able to analyze circuits in the new way. In relative terms, students taught the new way perform significantly better in analyzing circuits than do students in traditional courses.[10,11]

The main goal of this paper is to present a new approach for the teaching of circuits. The discussion parallels the treatment in our textbook,[1] but with additional background material of use to instructors.

Our model of electrons in metal is essentially classical, as is customary in introductory treatments of electricity and magnetism. In particular, electrons in metal conductors are treated according to the Drude model and are thought of as point particles, not delocalized as in the quantum description of metals. There are however echoes of this delocalization in the emphasis in our textbook on shifts of a mobile electron "sea" in a metal, both in electrostatics and in circuits. The Drude model is adequate to explain a broad range of phenomena in DC and RC circuits but of course is inadequate for explaining various quantum effects in metals.

In Section II we present an overview of the unified approach. In Section III we review the sizable literature on the subject.

# II. A UNIFIED APPROACH

## A. Preamble to the study of circuits

Before discussing the new approach to circuits, we provide some context by outlining the instruction that precedes the study of circuits. The first two chapters of our textbook deal with polarization phenomena in depth. Students perform simple "desktop" electrostatic experiments[12] in the lecture or recitation classroom that raise deep puzzles about important conceptual issues, such as the nature of static cling. These puzzles are then resolved in terms of atomic-level models of polarization in insulators, metals, and ionic solutions. The next two chapters introduce the concept of electric field, including the field of distributed charges (charged rods, rings, disks, parallel-plate capacitors, and spheres).

The fifth chapter is a mainly phenomenological introduction to circuits. Students do desktop circuit experiments which raise deep puzzles, such as why the battery doesn't always produce the same current. The biggest puzzle of all, which is discussed in the following section, is resolved in the sixth chapter, drawing heavily on the earlier work on electric field and on the atomic nature of matter.

## B. A puzzle concerning a simple circuit

We ask students to observe a simple circuit consisting of a battery with two conducting wires leading to a flashlight bulb in a socket (Figure 2). The current in the bulb filament is driven by an electric field inside the filament, with $\vec{J} = \sigma\vec{E}$ (for the introductory student we write this in the microscopic form i = nAv = nAuE, where i is the electron current in electrons per second, n is the free-electron density, A is the cross-sectional area, and the drift speed v = uE, where u is the electron mobility). We continually make the point that the electric field goes to zero inside a metal in static equilibrium, but that a circuit is a system which is kept out of static equilibrium, and the electric field inside a current-carrying wire is *not* zero.

It is natural to think that the source of the electric field in the bulb filament is charges in and on the battery. But if we bend the wires to bring the bulb closer to the battery, the bulb doesn't get brighter, which implies that the electric field inside the filament is unchanged despite the fact that the filament is much closer to the hypothetical source charges. Moreover, if we rotate the socket by 90 degrees (with the wires remaining attached), the electric field due to the battery charges should now be perpendicular to the filament, and there should be little or no current, yet we don't see any change in the brightness of the bulb.

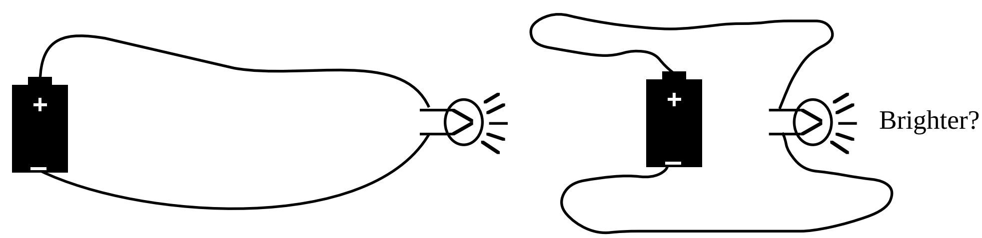


Figure 2. If the charges responsible for the electric field inside the bulb filament are in and on the battery, shouldn't the bulb be much brighter when brought closer to the battery?

Evidently charges in and on the battery *cannot* be the only source of the electric field inside the bulb filament. There is a deep puzzles: *Where are the charges that are the source of the electric field inside the bulb filament?*

One can try to avoid this question by assuming that the flowing electrons push each other through the wire, like water molecules pushing neighboring water molecules through a hose. However, this cannot be the explanation. Averaged over a few atomic diameters, the interior of the metal is everywhere neutral, and on average the repulsion between flowing electrons is canceled by attraction to positive atomic cores (Figure 3). This is one of the reasons why an analogy between electric current and the flow of water can be misleading.

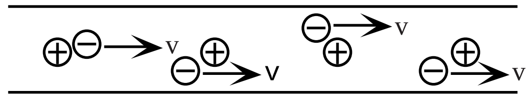


Figure 3. The flowing electrons inside a wire cannot push each other through the wire, because on average the repulsion by any electron is canceled by the attraction of a nearby positive atomic core.

Of course it is easy to understand the constant brightness in terms of potential: The potential drop across the bulb is always nearly equal to the emf of the battery, and the electric field inside the filament is simply the gradient of this constant potential drop. But we know that charges are the only sources of electric fields (in the absence of varying magnetic fields), and it is legitimate to ask where are the charges that are responsible for the field inside the filament.

If we don't ask and answer this question, we cannot hope to unify electrostatic and circuit phenomena, and we find ourselves forced to say rather lamely that in electrostatic situations, electric fields are made by charges, but in circuits, electric fields are made by potential differences. This is very unsatisfactory, since a major goal of introductory physics courses is to show how a small number of fundamental principles can explain a wide range of phenomena.

## C. Feedback and surface charge

So that we can ask students to account for all charges unambiguously, we ask students to consider a circuit driven by a charged capacitor instead of a chemical battery. The geometry of the circuit has been chosen to illustrate in a particularly dramatic fashion the relationship between charge and field in a circuit. We present a condensed version of the reasoning about this circuit that we ask students to perform. Because of its greater familiarity to physicists, we write $\vec{J} = \sigma\vec{E}$ for ohmic materials instead of the microscopic formula initially used by the students for electron current, i = nAuE.

The students are asked to consider a large-capacity parallel-plate capacitor which discharges through a high-resistance wire with some twists and turns (Figure 4), so that it takes a long time for the capacitor to discharge. For a considerable period of time (a fraction of the RC time constant), there is a quasi-steady state and a nearly constant conventional current I in the wire. (Alternatively, one can sustain a steady current without the complexities of a chemical battery by introducing a mechanical charge transport mechanism between the capacitor plates, similar to the conveyor belt in a Van de Graaff generator.) If the wire has a constant cross section, in the quasi-steady state the magnitude of the current density $\vec{J}$ and of the electric field $\vec{E} = \vec{J} / \sigma$ must be uniform throughout the wire.

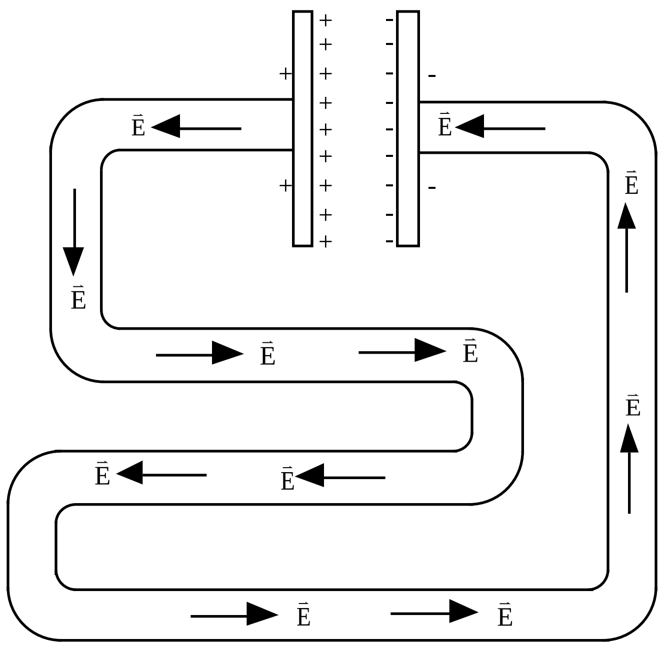


Figure 4. Discharging a large-capacity parallel-plate capacitor through a very thin resistive wire of constant cross section (not drawn to scale). In the quasi-steady state, the electric field inside the wire must be uniform in magnitude and parallel to the wire.

But the charges on the capacitor plates don't produce the required quasi-steady-state pattern of electric field (Figure 5). Not only do the capacitor charges fail to make a electric field with uniform magnitude, but the field $\vec{E}_{cap}$ due to the capacitor charges even points in the wrong direction between the left and right bends of the wire (and although not shown in the diagram, at many positions in the wire the electric field produced by the capacitor is not parallel to the wire). Evidently there must be some other charges somewhere, and the electric field of these unidentified charges plus the electric field of the capacitor charges must add up to make the quasi-steady-state uniform electric field shown in Figure 4. *Where are these other charges?*

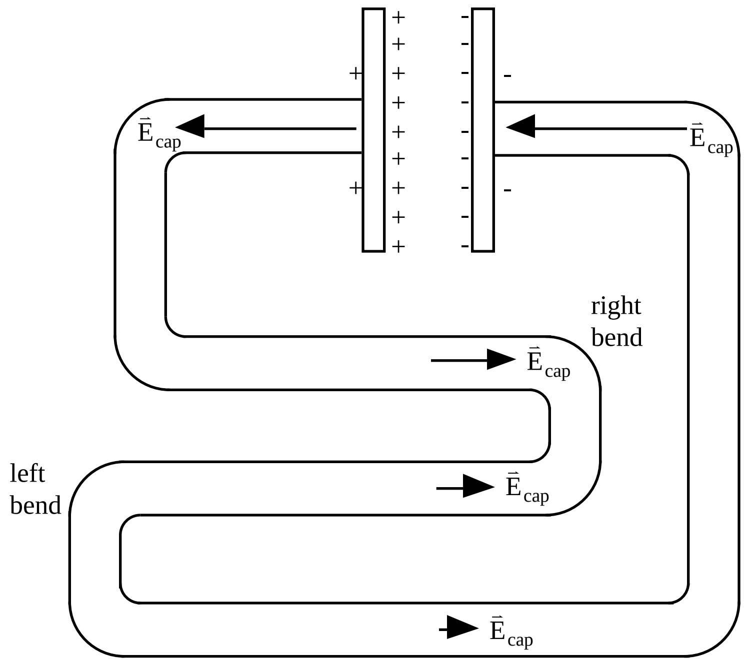


Figure 5. The electric field $\vec{E}_{cap}$ due to the charges on the capacitor plates is not uniform in magnitude, and between the left and right bends it even points in the wrong direction for driving the current in the quasi-steady state.

A thought experiment is useful. Imagine that we can attach the wire to the capacitor while preventing the mobile electrons in the wire from moving. Then we release the electrons and see what happens (Figure 6). Look at what happens in the region of the left bend: The initial electric field drives electrons into *both* ends of the left bend. This region therefore acquires a net negative charge, and this excess negative charge must reside on the surface of the metal. Similarly, the initial electric field drives electrons out of both ends of the right bend, leaving this region with a net positive charge, and this excess positive charge must be on the surface of the metal.

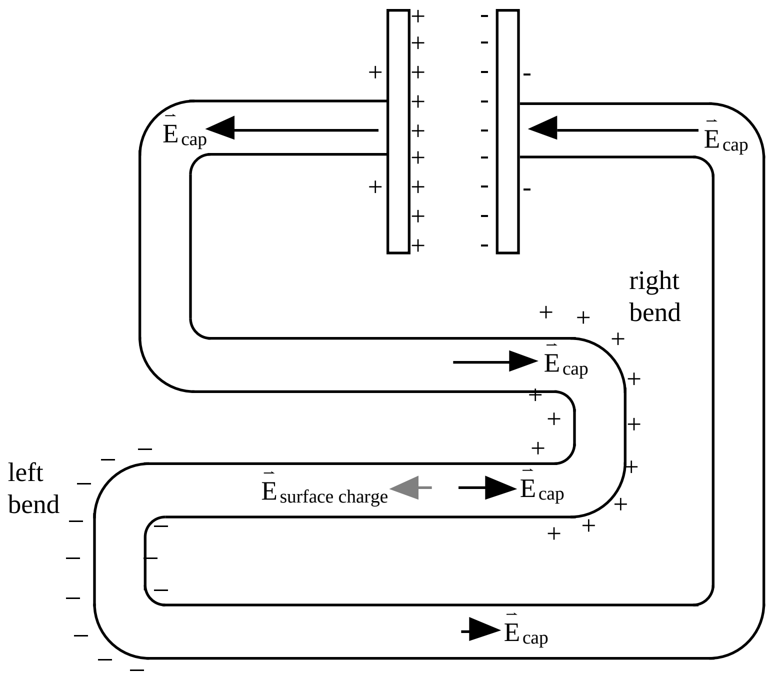


Figure 6. The initial electric field due to the charges on the capacitor drives electrons out of the right bend and into the left bend, leading to surface charge pile-up. The electric field of these new surface charges is to the left. The surface charge will continue to pile up until the *net* electric field is to the left.

After a very short time the non-steady-state configuration of electric field causes surface charges to appear on the wire. What contribution to the net electric field do these charges make? In the branch between the bends, where the electric field initially points the "wrong" way (to the right), the new surface charges make an electric field in the "correct" direction (to the left; see Figure 6). If there is insufficient surface charge on the bends to make a *net* electric field to the left, of the appropriate magnitude, pile-up of surface charge will continue until the net electric field does point to the left, with the appropriate quasi-steady-state magnitude.

Ultimately, this elegant negative feedback mechanism will produce an arbitrarily complex distribution of surface charge all over the wire such that the electric field due to *all* the charges (surface charge plus the charge on the capacitor plates) is uniform throughout the wire and has the magnitude predicted by conservation of energy (the Kirchhoff loop rule). The quasi-steady-state charge distribution might look *very* roughly like Figure 7, with a variation of surface charge density along the wire from the positive capacitor plate around to the negative capacitor plate, although the actual details of the surface-charge distribution for any particular circuit geometry may be quite complicated.

In instruction, we emphasize that the pattern of electric field is simple, while the distribution of surface charge is complex and extremely difficult to calculate accurately (see Section III for published calculations for a variety of geometries). In some simple geometries one can appeal to continuity arguments. In particular, the surface charge must be positive near the positive end of the battery and negative near the negative end of the battery. The pedagogical importance of the surface charge is that it provides a mechanism for understanding at a fundamental level why a circuit behaves as it does. It is however the pattern of electric field that can be drawn with confidence; any guess at the distribution of surface charge will be very approximate.

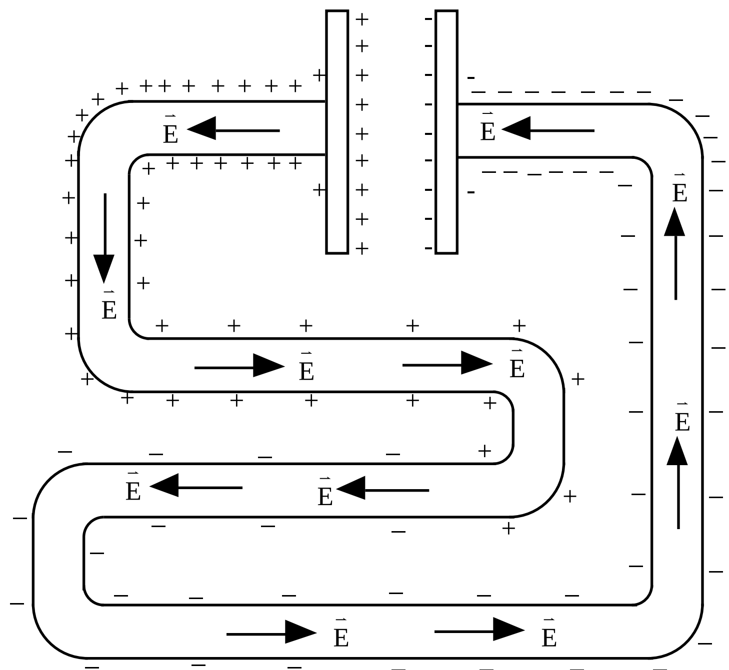

Figure 7. A possible, *very* approximate surface-charge distribution which, together with the charges on the capacitor plates, could produce the simple pattern of uniform-magnitude electric field.

We know experimentally that the appropriate surface charges are established very rapidly in a simple resistive circuit, because we do observe a steady state almost immediately after connecting the circuit or bending the wires. If there were no dissipation the feedback mechanism could overshoot and lead to oscillations rather than a steady state. (Haertel[13] has pointed out that the establishment of the steady state does not actually take place in one step, but rather is the end result of repeated reflections of waves sloshing around in the circuit before damping out due to dissipation.)

## D. Electric field of rings of surface charge

Our initial discussion focused on what happens on the bends in a particular circuit because the effect is so dramatic. But even on a straight section of wire there must be something like a *gradient* of surface-charge density to produce the required electric field inside the wire. In order to help students see that a non-uniform surface-charge distribution could produce a uniform electric field in a wire, we consider a concrete and simple case of rings of charge sheathing a section of wire, with a gradient of surface-charge density along the wire.

Before starting the study of circuits, our students study the electric field of distributed charges, including the on-axis electric field of a uniform ring. The axial field of a ring is useful in understanding the relationship between the surface charge on a wire, considered as a sequence of rings of charge, and the electric field inside the wire. We lead the students to see that a *gradient* of surface charge would contribute to an electric field inside the wire. Rings of equal charge density (and the same sign) contribute *zero* electric field at a location midway between the two rings, whereas rings of unequal charge density (or different sign) contribute a non-zero field at that location (Figure 8).

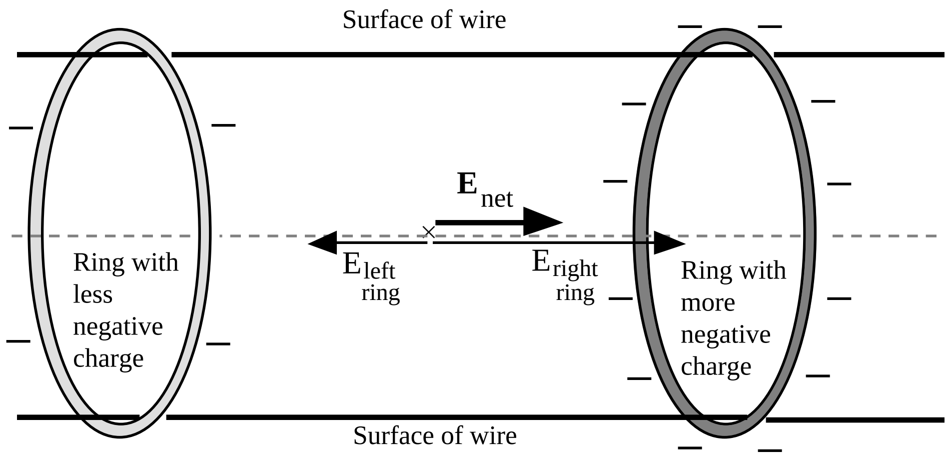


Figure 8. At a location midway between two surface-charge rings with different amounts of charge, the rings contribute a non-zero net electric field. Electric field inside the wire is associated with a *gradient* of surface-charge density.

In one particularly simple geometry, that of a long straight coaxial cable driven by a battery at one end and short-circuited at the other end, the uniform electric field inside the center conductor is actually produced by a constant gradient of surface charge density.[14,15] In other circuit geometries the distribution of surface charge may be arbitrarily complicated even though the pattern of electric field is simple.

Two rough approximations make it possible for students to reason qualitatively about surface-charge distributions and the mechanism for establishing the steady-state distribution of electric field, at least for the case of relatively simple circuit geometries. First, due to the inverse-square distance dependence of electric field, nearby surface charges may contribute a sizable fraction of the net electric field at a point (although of course *all* the surface charges everywhere in the circuit together determine the net field). This permits reasoning locally in a qualitative way about charge and field. Second, electric fields in lumped resistors may be thought of as being due in large part to gradients of surface charge along the resistor.

These are very rough approximations and applicable only for simple circuit geometries, but they are pedagogically useful because they allow students to reason qualitatively about circuit behavior, and to understand the basic mechanism that creates and maintains the steady state. In Section III we review the exact calculations of surface charge that have been carried out for a variety of geometries. These calculations confirm the approximate validity of the picture of surface charge sketched here.

We emphasize however that ultimately it is the relatively simple pattern of electric field that can be counted on, whereas the distribution of surface charge can be arbitrarily complex, especially if the circuit geometry is complicated.

## E. Lecture demonstration of surface charge

We have used a helpful demonstration of surface charge suggested by the science education group at the Weizmann Institute[16] (see Figure 9). A chain of four 80-megohm resistors is supported in air from

Styrofoam columns, so that the resistor chain is far from other objects, including the table, to avoid unwanted polarization effects. The resistor chain is connected to two DC power supplies which apply +5000 volts to one end of the resistor chain and -5000 volts to the other end, with respect to ground.

We model the resistors as very thin wires. From current conservation (I = enAv = enAuE, where u is the electron mobility), students can predict that the electric field must be very much larger inside the thin resistors than inside the thick connecting wires, and the electric field must have the same magnitude inside each resistor. As usual, the pattern of electric field is simple. The feedback mechanism will lead to an arbitrarily complicated distribution of surface charge that generates this simple pattern of electric field.

In Figure 9 we show a plausible rough, approximate distribution of surface charge, based in the first instance simply on continuity (positive near the positive pole of the power supply, negative near the negative pole). We then ask the students to check whether this estimated charge distribution is consistent with the pattern of electric field. Very roughly, the electric field inside the left-most resistor (AB) is in large part due to a large negative surface-charge density at A and a medium surface-charge density at B, with a large gradient along the resistor. The electric field must have the same magnitude in the next resistor (BC) and is largely due to the medium surface-charge density at B and the nearly zero surface-charge density at C. Similar remarks apply to the other two resistors. Inside the connecting wires the electric field is very small, and we indicate a roughly uniform charge distribution along these wires (the details of the actual charge distribution, except for the sign of the charge, are presumably more complicated).

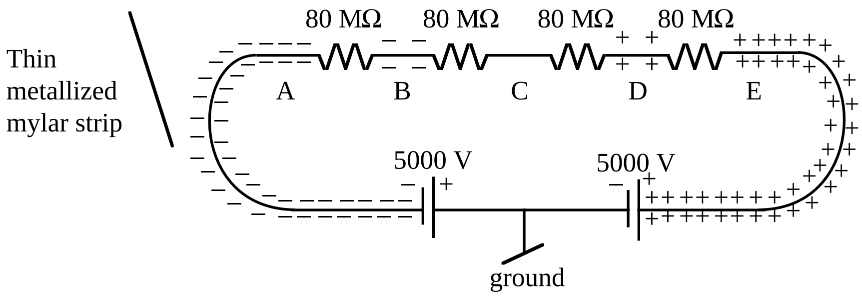


Figure 9. A lecture demonstration of surface charge. A thin metallized strip of mylar is attracted, charged by contact, and repelled by the surface charge. The distribution of surface charge is *very* approximate.

A flexible, thin metallized strip of mylar is suspended from an insulating rod and brought the bare wire near location A, where the strip is observed to be attracted to the charge-carrying wire (due to polarization of the strip) and then to jump away, due to being charged by contact and repelled by the surface charge at that location. The strip is found to be negatively charged, because it is repelled by a plastic pen rubbed through one's hair (a plastic pen is known to charge negatively). A similar effect is observed at E, but the strip is observed to be positively charged. A smaller effect is observed at B and D, where the surface-charge density is expected to be smaller. No effect is observed at point C, where by symmetry there is essentially no surface charge. These experimental observations confirm the validity of the approximate surface charge analysis.

The surface-charge density is proportional to the circuit voltage, and only at very high voltages is there enough charge to observe electrostatic repulsion in a mechanical system. It is difficult to observe repulsion by surface charge in a low-voltage circuit, because a charged object is of course *attracted* to neutral matter, and the attraction between a charged object and a circuit can mask the repulsion, if the surface charge on the circuit is small, as it is in low-voltage circuits. The object must be very light so that one can observe the interaction at a sizable distance, in order to minimize the competition between repulsion and the attraction due to polarization of neutral matter, which falls off with distance much more rapidly than $1/d^2$. (The attraction between a point charge and a small neutral object is proportional to $1/d^5$, since the polarization of the neutral object is proportional to $1/d^2$, and the interaction between a point charge and a dipole is proportional to $1/d^3$.)

# F. Quantitative circuit analysis

Here we review the steps leading to quantitative analysis of circuits in terms of electric field. Early in our study of circuits we introduce the important formula for electron current, $i = nAv$, where i is the number of electrons per unit time that pass some location in the circuit, n is the number of free electrons per unit volume, A is the cross-sectional area of the wire, and v is the electron drift speed. The students carry out experiments which show that in the steady state the current leaving a circuit element is equal to the current entering the element. We lead the students to see that if the outgoing electron current were smaller than the incoming current, the circuit element would become increasingly negatively charged, which would slow down the incoming electrons and speed up the outgoing electrons, until the two currents become equal. Current conservation is the result of charge conservation plus the dynamic feedback mechanism that establishes the steady state in a circuit.

Next we introduce the mobility u, a property of the material, with $v = uE$ (the drift speed is proportional to the applied field E). This permits the students to compare the magnitude of the electric field inside wires of different thicknesses that are in series. It is extremely useful that students do not recognize the formula $v = uE$ as being just a microscopic version of Ohm's law ($I = \Delta V/R$). In our experience, students who have learned Ohm's law in a previous course typically over-generalize this approximate description of material properties, applying it inappropriately to batteries as well as resistors, and even confusing Ohm's law with the much more fundamental Kirchhoff loop rule.

For example, many students think that there cannot be a potential difference across an open switch, because $V = IR$, and if there is no I there cannot be any V. An even more extreme example of an unthinking over-reliance on Ohm's law was provided by a student who agreed on experimental and theoretical grounds that the current leaving a light bulb has to be equal to the current entering the bulb, but argued that the current in the bulb filament itself has to be much less, since $V = IR$, and the bulb filament has more R than the neighboring wires so it must have less I! Emphasis on a microscopic view, and the relation $v = uE$, gives the students a strong microscopic sense of mechanism and reduces unthinking rote application of $V = IR$.

We introduce a "mechanical battery" similar to a Van de Graaff generator, in which a conveyor belt applies a non-Coulomb force $F_{NC}$ to extract an electron from the positive plate of a large parallel-plate capacitor and push it onto the negative plate (Figure 10). In an open circuit, or with small currents (and low internal resistance), $F_{NC} = eE_C$, where $E_C$ is the Coulomb field due to the charges on the plates. Such a mechanical battery can be understood in much greater detail than a chemical battery, and it provides a model for a device that maintains a charge separation (and later helps in explaining the important distinction between emf and potential difference).

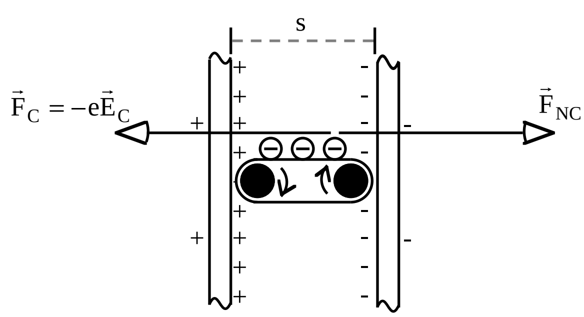


Figure 10. An isolated "mechanical battery." Charge builds up on the plates until the Coulomb force $eE_C$ is equal to the non-Coulomb force $F_{NC}$. The separation s is small compared to the size of the plates.

In a circuit containing a mechanical battery and a series of wires, we show that energy conservation leads to an equation relating the amount of work done on an electron by the mechanical battery to the energy dissipated in the wires, where $E_i$ is the electric field in a wire of length $L_i$:

$$F_{NC}s = \sum eE_iL_i$$

At this point students can analyze simple circuits in terms of a microscopic picture, using this energy-conservation equation plus the equation for electron-current conservation (the electron current $i = nAv =$

nAuE is the same in series circuit elements). Figure 11 gives an example of a typical homework problem. The student is asked to draw the electric field at the locations marked ×, (small inside the thick wires, large inside the thin wire) and to sketch an approximate surface charge distribution (positive along the left branch, negative along the right branch, with a strong gradient of surface charge along the thin wire). Then the student is asked to calculate quantitatively the number of electrons entering the thin wire every second.

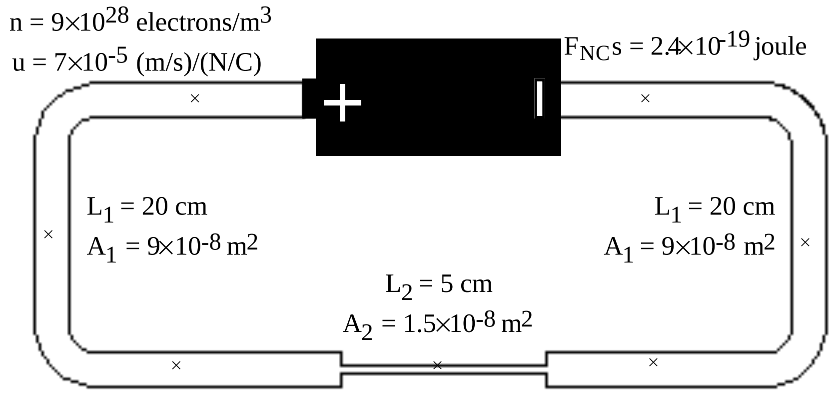


Figure 11. An example of a typical homework problem in which a circuit is analyzed quantitatively in terms of charge and electric field. The thick and thin wires are made of the same resistive metal with given mobility u, density n of free electrons, lengths, and cross-sectional areas. The non-Coulomb work per unit charge $F_{NC}s$ corresponds to a 1.5 volt battery.

It is not possible with elementary techniques to determine the exact form of the surface-charge distribution. What is calculable from current conservation and energy conservation is the electric field. From the pattern of electric field, one can sketch an approximate surface-charge distribution for simple circuit geometries such as the one shown in Figure 11.

The mobility u is a property of a material, but it is in general temperature-dependent. We have the students carry out a desktop experiment with flashlight bulbs in which they find that the current in a two-bulb series circuit is significantly larger than half the current found in a one-bulb circuit. One can show from the energy-conservation equation that the electric field in a bulb filament is exactly half as big in the two-bulb circuit (assuming very small fields in the connecting wires), so one concludes that in the equation $i = nAuE$ the mobility is greater, due to the lower temperature of the dimmer bulbs.

In this microscopic framework, with emphasis on charge and field, some very common student misconceptions melt away. One common misconception is that the potential difference across an open switch is zero (because V = IR, and if I = 0, V must be zero). Our students see that the two branches of the circuit leading from the battery to the open switch carry positive and negative surface charge, and these charges make an electric field in the gap, with an associated potential difference. This comes up after we introduce the concept of potential as a path integral of electric field and then re-analyze circuits in terms of potential, but without forgetting about surface charge and electric field.

Another common misconception is that current is used up in a light bulb. Our students have a strong sense of mechanism and have a clear microscopic mental model of electrons drifting through the bulb filament, pushed by the electric field of the surface charges, and that current conservation is the result of the dynamic feedback mechanism. They are familiar with the steady state being established by buildup of surface charges, which focuses attention on the fundamental mechanisms as opposed to reliance on what are for the students vague, non-mechanistic notions of current and potential.

Students often think that a battery either outputs zero current (if nothing is attached to it) or outputs a standard amount of current, independent of what is attached to the battery. Again, a strong sense of microscopic mechanism, and the use of a “mechanical battery” in various explanations, helps combat this deep-seated misconception of the role of batteries in circuits.

Analyses of a variety of simple DC and RC circuits in terms of charge and field can be found in chapters 6 and 7 of our textbook.[1] One of the important discussion topics in chapter 6 is the nature of the transient that leads to establishing the steady state. It is often said to be a common student "misconception" that a bulb will light when attached by a wire to only one end of a battery. In our dynamic view of circuits, this is not entirely a misconception. As the wire is brought toward the battery, a tiny transient current does flow, polarizing and charging the wire and bulb. It is only a quantitative matter that the current is so small and lasts such a short time that one doesn't see the bulb light up. In fact, in Steinberg's experiments[17-20] and in the chapter on capacitor circuits in chapter 7 of our textbook, the gap in the circuit may be a one-farad capacitor, and the bulb shines brightly for a time without a closed circuit. We see here a significant advantage to treating electrostatics and circuits in a unified way. Also, emphasis is placed on dynamic processes rather than solely on the steady state, an emphasis seen in both Haertel's and Steinberg's work.

## G. Student questions

The analysis of circuits in terms of surface charge provides answers to some profound questions students sometimes raise, questions which may have no satisfactory answer within the context of traditional analysis in terms solely of potential. Haertel[5] reports students asking, "What is the difference in the conditions at the two ends of a resistor?" It is quite unsatisfying to say merely that the potential is different, and quite satisfying for the student to see that a difference in surface-charge density at the two ends of the resistor leads to a large electric field inside the resistor, which drives the current. Another student question reported by Haertel is, "How does the current know how to split when there are parallel branches?" During the initial transient phase, current may run equally down two parallel branches, but different resistance in the two branches leads to different surface-charge buildup along the two branches, which in the steady state will steer appropriate amounts of current down each branch.

## H. Experience with teaching surface-charge analysis

We have found that surface-charge analysis can be made teachable and learnable, and that it provides a strong sense of mechanism as well as unifying electrostatics and circuits. The aspect that seems to pose the greatest difficulty for students is the distinction between surface-charge density and its gradient. There is a not unnatural tendency to think that a large *amount* of surface charge rather than a large *gradient* of surface charge implies a large electric field. This is yet another example of the difficulty students have throughout introductory physics in distinguishing between quantities and rates.

The advantages and disadvantages of the new approach compared with the traditional approach to circuits are quite similar to those of kinetic theory compared with thermodynamics. Kinetic theory offers a detailed, mechanistic picture which provides a basis for a qualitative physical understanding of how the steady state arises, whereas thermodynamics typically starts from the steady state, which can seem mysterious and unphysical, since it may not be clear how this steady state would arise and maintain itself. The disadvantages shared by the new electricity approach and kinetic theory are that situations where exact numerical results for all aspects of the phenomena cannot be easily obtained, and in many cases even qualitative conclusions about system behavior may require more detailed and lengthy analysis than when using more powerful and more abstract principles such as potential or thermodynamics.

### Analysis in terms of potential

After analyzing circuits in terms of surface charge and electric field, in our textbook we introduce electric potential difference as a path integral of the electric field. Electric field is considered to be the more fundamental concept in this introductory course, and potential difference is considered a secondary concept based on electric field. In the chapter on potential, both electrostatic and circuit patterns of electric field are used to illustrate potential difference.

The chapter on potential is followed by a chapter devoted entirely to potential and circuits. The loop rule however is approached not as being special to circuits but rather as a consequence of the general principle that the round-trip integral of electric field must be zero, if the electric field is due to (stationary) point charges. In a circuit with a "mechanical battery" the students can see that the Coulomb electric field inside the “battery” points in the opposite direction to the conventional current flow, and the round-trip path integral of electric field is zero.

In this circuit chapter we distinguish very carefully between emf (non-Coulomb work per unit charge, as in the “mechanical battery” where emf = $F_{NC}s/e$) and potential difference (path integral of the Coulomb electric field). When we study magnetic induction later in the course, we put more emphasis than usual on the non-Coulomb electric field associated with a time-varying magnetic field, and this too helps distinguish between emf and potential difference.

The new approach provides additional useful practice with the relationship between charge and field. When circuits are analyzed solely in terms of potential, one loses an important opportunity to deepen students’ understanding of electric field by additional practice with the concept in a novel setting. Even when we do analyze circuits in terms of potential, we don’t drop electric field but continually make connections between the electric field inside the wires and resistors and the potential differences along those elements. This continuing emphasis on electric field helps students see that it is potential difference, not potential, that is physically meaningful.

By analyzing electric phenomena first in terms of charge and field, and only later in terms of potential, we are paralleling the usual sequence in the mechanics course, where force is introduced and used before energy methods are introduced. Analysis in terms of force or field seems more concrete and less abstract but is often less powerful than analysis in terms of energy or potential.

It has been common to use hydrodynamic or other analogies in the teaching of electricity, but these analogies can be misleading. Another advantage of the new approach is that it is based directly on the fundamental Coulomb interaction rather than on a necessarily inaccurate analogy to some other physical process.

In addition to calling for treating fewer topics in more depth, the Introductory University Physics Project (IUPP) has called for having introductory physics courses place more emphasis on 20th-century physics topics.[21,22] Our new approach to the teaching of electricity addresses the latter concern in two significant ways. We have placed more emphasis than is usual on the key role of electrons in metals, and on atomic-level polarization phenomena in insulators and metals. Also, as Haertel[23] has pointed out, an electric circuit analyzed in terms of surface charge, while richly complex, is nevertheless one of the simplest systems that fully manifests the phenomena of an interactive *system* in the modern sense, including feedback and the interplay between local and global aspects.

# III. A REVIEW OF THE LITERATURE

In this section we summarize important points in the literature that are particularly relevant to teaching about the role of electric field in circuits. We also provide a historical note that provides additional perspective on the status of this analysis.

## A. Small charges have big effects

Inside a resistive wire 1 meter long connected to a 3 volt battery, the electric field is only 3 volts/meter, which is tiny compared with typical fields encountered in electrostatic phenomena (for example, the breakdown strength of air is about 3 *million* volts/meter). Therefore, the amount of surface charge on the wires of a typical circuit is extremely small compared to typical electrostatic charges, which is why it requires high-voltage circuits to observe electrostatic effects. However, these small amounts of charge are responsible for driving the current inside the wire. Applying the quantitative analyses of Sommerfeld[14] and Marcus[15] to a straight coaxial cable 30 cm long whose central conductor has a diameter of 1 mm, driven by a 3 volt battery at one end and shorted at the other end, the maximum linear surface charge density on the central conductor is only a few million electrons per centimeter.

Parker[24] points out that the surface charge on the wires of a circuit is maintained by a dynamic equilibrium. If a fluctuation should lead to there being more positive surface charge at some place than is required for steady-state current flow, that portion of the surface will attract electrons, which will reduce the imbalance. Parker also offers a vivid picture of what happens if a straight current-carrying wire is bent: Electrons pile up on the outer surface of the bend until there is enough negative charge there to turn the oncoming electrons around the bend. Rosser[25] calculates that a single electron on the surface of a right-angled bend in a copper wire is sufficient to turn an ampere of current, about $10^{19}$ electrons per second! Haertel comments, “Because of the enormous strength of the Coulomb interaction and the very high mobility of electrons in metals, it takes only a few electrons at the surface of the wire to push $10^{19}$ electrons around in a circle and to overcome the resistance of a metallic wire” (Ref. 5, p. 42).

## B. Contributions of nearby surface charges

In some simple cases the dominant contribution to the electric field inside a wire may be made primarily by nearby surface charges, because the $1/d^2$ character of the Coulomb interaction may make the contributions of distant charges quite small. To the extent that we can neglect the effects of distant surface charges, we can make qualitative analyses in terms of local causes and local effects.

In Figure 12 is sketched an analysis by Walz[26] of the relative importance of nearby and far-away charges when there is a constant gradient of surface charge along a straight wire. One finds that the peak contribution dE to the field comes from surface charges at a distance of 1.4r, from which Walz concluded that E is due mainly to nearby charges.

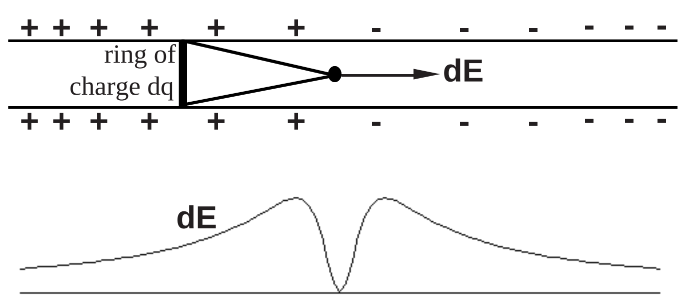


Figure 12. According to Walz, most of the contributions dE to the field E come from nearby rings of surface charge (constant gradient).

However, with a constant gradient of surface-charge density (dq proportional to xdx) a distant ring contributes an amount dE proportional to $(xdx)/x^2 = dx/x$, so the contribution of distant charges along an infinite straight wire diverges logarithmically. In a real circuit geometry this logarithmic divergence will be cut off due to bends in the wire. Walz's argument is suggestive that nearby charges may make a major contribution to the electric field in simple geometries, but it is easy to find examples where this approximation breaks down. For example, if a wire crosses itself or there are multiple loops as in a coil, longitudinally remote but geometrically nearby sections of the wire make large contributions to the field.

## C. Calculations and measurements of surface charge

Here is a summary of the most relevant theoretical and experimental literature on the role of surface charge in circuits.

Exact calculations of the distribution of surface charge are quite difficult, and such calculations as do exist deal with rather special circuit geometries. Schaefer[27] seems to have been the first to publish a quantitative analysis of the surface charge along a wire. Sommerfeld[14] gave a detailed analysis of a straight coaxial cable driven by a battery at one end and shorted at the other, and a similar but less complete treatment was carried out by Marcus.[15] Additional commentary is given by Russell.[28] Varney and Fisher[29] review these early calculations. Heald[30] treated an infinitely long circular cylindrical resistive circuit with azimuthal current and an infinitely long circular cylindrical circuit consisting of perfect conductors and a resistor. Saslow[31] calculated the properties of a spherical battery embedded in a conducting medium, including the surface charge on the battery. Aguirregabiria, Hernández, and Rivas calculated the surface charge on a square circuit driven by changing magnetic flux,[32] and the surface charge on a conducting ring rotating in a magnetic field.[33] Jackson[34] calculated the surface charge and field for a finite-length coaxial cable with conducting end plates, with a battery and lumped resistor placed at various locations along the cable.

Jefimenko[35,36] not only analyzed a number of interesting circuit configurations but built an ingenious demonstration device to make visible the electric fields surrounding a circuit. A similar device was reinvented by Parker,[24] and a different kind of demonstration apparatus was designed by Moreau, Ryan, Beuzenberg, and Syme.[37] Moreover, Parker[24] offered excellent physical insights into the role of surface charge and the nature of the feedback mechanism. Rosser[25,38] also provided useful physical insights into the mechanism of current flow in circuits. Moreau[39] gave an excellent discussion of the conservative and non-conservative fields in a circuit, and made brief but fruitful reference to the surface charge.

In Haertel's lengthy monograph[5] there are excellent physical insights and a major concern for the related pedagogical issues. In particular, he argues that surface charge isn't an exotic and peripheral aspect of circuits but provides an essential mechanism for understanding cause and effect. An earlier paper by Haertel[4] contains a brief version of these arguments, and in the same conference proceedings may be found the analysis by Walz discussed earlier.[26]

A group at the University of Thessaloniki, Greece, has been experimenting with teaching some elements of surface-charge analysis to younger students.[40] They have stressed the pedagogical importance of dealing explicitly with battery life under load, in order to incorporate energy considerations and all of the non-steady-state aspects of simple circuits.

## E. Historical note on surface charge

A doctoral thesis by Benseghir[41] contains an interesting and valuable historical summary of the development of circuit concepts, starting with Volta's invention of the Voltaic pile (battery) in 1800 and culminating with Kirchhoff's synthesis in 1849. For many physicists who had been studying electrostatics, the Voltaic pile was interesting only in terms of its electrostatic properties in an open circuit, especially because a closed circuit seemed to exhibit no electrostatic manifestations. It was thought that electrostatic considerations, and what we would call potential, did not apply to a closed circuit.

In 1820 Ampère partially sorted out the relationship between electrostatic and current phenomena.[42] While establishing Ohm's law for conductors in 1827, Ohm made a major contribution to circuit theory that went well beyond this relationship, because he clarified the separate and complementary roles of current and potential at a time when both were rather fuzzy concepts.[43] Moreover, Ohm conceived of there being two kinds of electricity in a closed circuit—a stationary gradient of volume charge (an error later corrected by Kirchhoff) corresponding to a gradient of potential, and a steady flow of electricity impelled by the stationary charge distribution, by analogy with a temperature gradient driving heat transfer. Ohm's work went unappreciated for about twenty years, because of its novelty.[44] French research from 1800 to 1830 that bears on these matters is discussed in a paper by Brown.[45]

Kirchhoff resolved the remaining fundamental issues, not just with his loop and node rules but more importantly by unifying electrostatics and circuits. Ohm's steady-state stationary charge distribution consisted of a gradient of volume charge inside the conductor, whereas on the basis of what was known about electrostatics there could not be a volume charge inside a conductor. Kirchhoff put the driving charge gradient on the surface of the conductors and showed that electrostatic and circuit phenomena belonged to one science, not two.[46] Unfortunately, this unification was later lost sight of as the role of potential came to dominate the analysis of circuits, and the surface charges disappeared from view.

# ACKNOWLEDGMENTS

We acknowledge a large debt to Hermann Haertel for teaching us about surface-charge analysis. We thank our colleagues Bat-Sheva Eylon, Jill Larkin, Fred Reif, and Raymond Sorensen for helpful discussions. We also thank an anonymous AJP reviewer who made thoughtful comments on a very early version of this article. This work was supported in part by NSF grant MDR-8953367 and NSF grant USE-9156105.

### References

[1] R. W. Chabay and B. A. Sherwood, *Electric and Magnetic Interactions* (Wiley, New York, 1995). An accompanying Instructor's Manual by the authors is available. For additional information see http://cil.andrew.cmu.edu/emi.html, which includes an extensive extract of the Instructor's Manual.

[2] B. A. Sherwood and R. W. Chabay, "Electrical interactions and the atomic structure of matter: adding qualitative reasoning to a calculus-based electricity and magnetism course," in *Learning Electricity and Electronics with Advanced Educational Technology*, NATO ASI Series F, Vol. 115, edited by M. Caillot (Springer-Verlag, Berlin, 1993), pp. 23-35.

[3] R. W. Chabay and B. A. Sherwood, "A new approach to introductory electricity and magnetism," forthcoming.

[4] H. Haertel, "The electric voltage," in *Aspects of understanding electricity: Proceedings of an international conference*, edited by R. Duit, W. Jung, and C. von Rhöneck (IPN/Schmidt & Klaunig, Kiel, Germany, 1985), pp. 353-362.

[5] H. Haertel, *A qualitative approach to electricity*, Report # IRL87-0001 (Institute for Research on Learning Palo Alto, CA, 1987).

[6] R. Cohen, B. Eylon, and U. Ganiel, "Potential difference and current in simple electric circuits: A study of students' concepts," Am. J. Phys. **51**, 407-412 (1983).

[7] L. C. McDermott and P. S. Shaffer, "Research as a guide for curriculum development: An example from introductory electricity. Part I: Investigation of student understanding." Am. J. Phys. **60**, 994-1003 (1992).

[8] J. L. Closset, "Sequential reasoning in electricity," in *Research on Physics Education: Proceedings of the First International Workshop* (Editions du CNRS, Paris, 1983), pp. 313-320.

[9] Shipstone, D. M. (1984). A study of children's understanding of electricity in simple D.C. circuits. European Journal of Science Education **6**, 185-198.

[10] B. Thatcher and D. Boys, "Student interpretation of transisents in DC electric circuits," AAPT Announcer **95**, no. 4, p. 95 (Dec. 1994). Details on one of the key comparison questions can be found at http://cil.andrew.cmu.edu/emi.html.

[11] Some early evaluation information is found on pages 9-10 of our Instructor's Manual (see reference 1). Additional information is available at http://cil.andrew.cmu.edu/emi.html.

12 B. A. Sherwood and R. W. Chabay, "Integrating theory and experiment in lecture using desktop experiments," to be published in the proceedings of the International Conference on Undergraduate Physics Education, University of Maryland, August 1966.

[13] H. Haertel, private communication, June 1991.

[14] A. Sommerfeld, *Electrodynamics* (Academic Press, New York, 1952), 125-130.

[15] A. Marcus, "The electric field associated with a steady current in long cylindrical conductor," Am. J. Phys. **9**, 225-226 (1941).

[16] U. Ganiel, private communication, January 1992.

[17] M. S. Steinberg, "Transient lamp lighting with high-tech capacitors," The Physics Teacher **25**, 95-96 (1987). We used NEC "Supercap" model FA0H474Z. This particular half-farad capacitor has a low "equivalent series resistance" (ESR) of 3.5 ohms, which allows it to be fully discharged quickly when shorted. Similar capacitors with larger ESR should be avoided because it is easy for students to become confused when a supposedly discharged capacitor actually holds a significant charge. There is a newer FE series which has an even lower ESR rating of 1.8 ohms.

[18] M. S. Steinberg, "Transient electrical processes as resources for causal reasoning," in *Proceedings of the 2nd International Seminar on Misconceptions and Educational Strategies in Science & Mathematics, Vol. 3* (Cornell University, Ithaca NY, 1987), pp. 480-490.

[19] M. S. Steinberg, *Electricity Visualized: The CASTLE Project* (PASCO Scientific, Roseville CA, 1992).

[20] M. S. Steinberg and C. L. Wainwright, "Using models to teach electricity—The CASTLE Project." The Physics Teacher **31**, 353-357 (1993).

[21] Introductory University Physics Project, Newsletter Number One, American Institute of Physics (1987).

[22] J. Rigden, D. Holcomb, and R. DiStefano, "The introductory university physics project," Physics Today 32-37 (April 1993).

[23] H. Haertel, private communication (June 1991).

[24] S. Parker, "Electrostatics and current flow," Am. J. Phys. **38**, 720-723 (1970).

[25] W. G. V. Rosser, "Magnitudes of surface charge distributions associated with electric current flow," Am. J. Phys. **38**, 265-266 (1970).

[26] A. Walz, "Fields that accompany currents," in *Aspects of understanding electricity: Proceedings of an international conference*, edited by R. Duit, W. Jung, and C. von Rhöneck (IPN/Schmidt & Klaunig, Kiel, Germany, 1985), pp. 403-412.

[27] C. Schaefer, *Einführung in die theoretische Physik* (de Gruyter, Berlin, 1932), volume 3, part 1, pp. 175-184.

[28] B. R. Russell, "Surface charges on conductors carrying steady currents," Am. J. Phys. **36**, 527-529 (1968).

[29] R. N. Varney and L. H. Fisher, "Electric field associated with stationary currents," Am. J. Phys. **52**, 1097-1099 (1984).

[30] M. A. Heald, "Electric fields and charges in elementary circuits," Am. J. Phys. **52**, 522-526 (1984).

[31] W. M. Saslow, "Consider a spherical battery...," Am. J. Phys. **62**, 495-501 (1994).

[32] J. M. Aguirregabiria, A. Hernández, and M. Rivas, "An example of surface charge distribution on conductors carrying steady currents," Am. J. Phys. **60**, 138-141 (1992).

[33] J. M. Aguirregabiria, A. Hernández, and M. Rivas, "An example of surface charge distribution on conductors carrying steady currents," Am. J. Phys. **64**, 892-895 (1996).

[34] J. D. Jackson, “Surface charges on circuit wires and resistors play three roles,” Am. J. Phys. **64**, 855-870 (1996).

[35] O. Jefimenko, “Demonstration of the electric fields of current-carrying conductors,” Am. J. Phys. 30, 19-21 (1962).

[36] O. Jefimenko, *Electricity and Magnetism* (Appleton-Century-Crofts, New York, 1966), pp. 299-304. A second edition was published by Electret Scientific Co., Star City, West Virginia (1989).

[37] W. R. Moreau, S. G. Ryan, S. J. Beuzenberg, and R. W. G. Syme, “Charge density in circuits,” Am. J. Phys. **53**, 552-553 (1985).

[38] W. G. V. Rosser, “What makes an electric current ‘flow’,” Am. J. Phys. **31**, 884-885 (1963).

[39] W. R. Moreau, “Charge distributions on DC circuits and Kirchhoff’s laws,” European Journal of Physics **10**, 286-290 (1989).

[40] D. Psillos, P. Koumaras, and A. Tiberghien, “Voltage presented as a primary concept in an introductory teaching sequence on DC circuits,” International Journal of Science Education, **10**, 29-43 (1988).

[41] A. Benseghir, *Transition électrostatique-électrocinetique: point de vue historique et analyse des difficultés des élèves* (Doctoral thesis, University of Paris VII, 1989).

[42] A. M. Ampère, “De l’action exercée sur un courant électrique par un autre courant,” Ann. Chem. Phys. **15**, (1820). Reprinted in *Mém. sur l’électromagnétisme et l’électrodynamique* (Gauthier-Villars et Cie, Paris, 1921). Extract in English translation in Ref. 39.

[43] G. S. Ohm, *Die galvanische Kette, mathematisch bearbeitet* (T. H. Riemann, Berlin, 1827). Included in *Gesammelte Abhandlungen von G. S. Ohm*, edited by E. Lommel (A. Meiner, Leipzig, 1892). French translation by J. M. Gaugain, *Théorie mathématique des courants électriques* (1860).

[44] M. Schagrin, “Resistance to Ohm’s law,” Am. J. Phys. **31**, 536-547 (1963).

[45] T. M. Brown, “The electric current in early nineteenth-century French physics,” Historical Studies in the Physical Sciences **1**, 61-103 (1969).

[46] G. Kirchhoff, “Ueber eine Ableitung der Ohm’schen Gesetze, welche sich an die Theorie der Electrostatik anschliesst,” Annalen der Physik und Chemie **78**, 506-513 (1849). English translation: “On a deduction of Ohm’s laws, in connexion with the theory of electrostatics,” Philosophical Magazine **37**, 463-468 (1850).