INTRODUCTORY

PHYSICS

Building Models to Describe Our World

Ryan Martin • Emma Neary • Olivia Woodman

License

Preface

About this textbook

This textbook is written to fill several needs that we believe were not already met by the many existing introductory physics textbooks. First, we wanted to ensure that the textbook is free to use for students and professors. Second, we wanted to design a textbook that is mindful of the new pedagogies being used in introductory physics, by writing it in a way that is adapted to a flipped-classroom approach where students complete readings, think about the readings, and then discuss the material in class. Third, we wanted to create a textbook that also addresses the experimental aspect of physics, by proposing experiments to be conducted at home or in the lab, as well as providing guidelines for designing experiments and reporting on experimental results. Finally, we wanted to create a textbook that is a sort of "living document", that professors can edit and re-mix for their own needs, and to which students can contribute material as well. The textbook is hosted on GitHub, which allows anyone to make suggestions, point out issues and mistakes, and contribute material.

This textbook is meant to be paired with the accompanying "Question Library", which contains many practice problems, many of which were contributed by students.

This textbook would not have been possible without the support of Queen's University and the Department of Physics, Engineering Physics & Astronomy at Queen's University, as well as the many helpful discussions with the students, technicians and professors at Queen's University.

Hello from the authors

Ryan Martin I am a professor of physics at Queen's University. My main research is in the field of particle astrophysics, particularly in studying the properties of neutrinos. I grew up in Switzerland, obtained my Bachelor's, Master's and Ph.D. at Queen's University. I was then a postdoctoral fellow at Lawrence Berkeley National Laboratory, a faculty at the University of South Dakota, before returning to Queen's. I am particularly passionate about education, and I am always seeking opportunities to involve students in helping to make education more accessible. I also like to cook and to play volleyball.

Emma Neary I am currently a second year physics major and QuARMS (Queen's University Accelerated Route to Medical School) student, as well as a native of St. John's, Newfoundland. Uniting the perspectives of students and professors in an accessible way is important to me. I strongly believe in the importance of building physical models; whether it be in physics, medicine, sciences or the arts. It has been my goal to infuse the textbook with the theme of modelling in a creative and engaging way. Aside from doing physics, I enjoy hiking, dancing, reading and doing research in gastroenterology and neuropsychiatry.

Olivia Woodman I am a currently a third year undergraduate student at Queen's Univeristy, majoring in physics. The flipped classroom approach has been beneficial to my own learning, and I think that we have created a textbook that really complements this learning style. Throughout this book, I have shared my thoughts on various topics in physics, as well as some useful tips and tricks. I hope that students enjoy using this book and continue to contribute to it in the future. Working on this textbook has also allowed me to combine my love of physics with my love of doodling, so I hope you enjoy the drawings!

How to use this textbook

This textbook is designed to be used in a flipped-classroom approach, where students complete readings at home, and the material is then discussed in class. The material is thus presented fairly succinctly, and contains **Checkpoint Questions** throughout that are meant to be answered as the students complete the reading. We suggest including these Checkpoint Questions as part of a quiz in a reading assignment (marked based on completion, not correctness), and then using these questions as a starting point for discussions in class.

For topics that are particularly difficult, we have included **Thought Boxes** written by students that try to present the material in a different light. We are always happy if students (or professors) wish to contribute additional thought boxes.

Chapters start with a set of **Learning outcomes** and an **Opening question** to help students have a sense of the chapter contents. The chapters have **Examples** throughout, as well as additional practice problems at the end. The **Question Library** should be consulted for additional practice problems. At the end of the chapter, a **Summary** presents the key points from the chapter. We suggest that students carefully read the summaries to make sure that they understand the contents of the chapter (and potentially identify, before reading the chapter, if the content is review to them). At the end of the chapters, we also present a section to **Think about the material**. This includes questions that can be assigned in reading assignments to research applications of the material or historical context. The thinking about the material section also includes experiments that can be done at home (as part of the reading assignment) or in the lab.

Appendices cover the main background in mathematics (Calculus and Vectors), as well as present an introduction to programming in python, which we feel is a useful skill to have in science. There is also an Appendix that is intended to guide work in the lab, by providing examples of how to write experimental proposals and reports, as well as guidelines for reviewing proposals and reports. We believe that introductory laboratories should not be be "recipe-based", but rather that students should take an approach similar to that of a researcher in designing (proposing) an experiment, conducting it, and reviewing the proposals and results of their peers.

Credits

This textbook, and especially the many questions in the Question Library would not have been possible without the many contributions from students, teaching assistants and other professors. Below is a list of the people that have contributed material that have made this textbook and Question Library possible.

Adam McCaw
Ali Pirhadi
Alexis Brossard
Amy Van Nest
Ceaira Heimstra
Damara Gagnier
Daniel Barake
Daniel Tazbaz
David Cutler
Emily Darling
Emily Mendelson
Emily Wener
Emma Lanciault
Genevieve Fawcett
Gregory Love
Haoyuan Wang
Jack Fitzgerald
James Godfrey
Jenna Vanker
Jesse Fu
Jesse Simmons
Jessica Grennan
Joanna Fu
Jonathan Abott
Josh Rinaldo
Kate Fenwick
Madison Facchini
Marie Vidal
Matt Routliffe
Maya Gibb
Nicholas Everton
Nick Brown
Nicole Gaul
Noah Rowe
Olivia Bouaban
Patrick Singal
Qiqi Zhang
Quentin Sanders
Robin Joshi
Ryan Underwood
Sam Connolly
Sara Stephens
Shona Birkett
Stephanie Ciccone
Talia Castillo
Tamy Puniani
Thomas Faour
Troy Allen
Wei Zhuolin
Yumian Chen
Zifeng Chen
Zoe Macmillan

Contents

1 The Scientific Method and Physics

Learning Objectives

- Understand the Scientific Method.
- Define the scope of Physics.
- Understand the difference between theory and model.
- Have a sense of how a physicist thinks.

Think About It

A scientific theory...

A) must explain the physical world, and it may or may not be experimentally verifiable.
B) proves our models to be correct, and it must be experimentally verifiable.
C) describes the physical world, and must be experimentally verifiable.
D) must disprove other theories, and may or may not be experimentally verifiable.

1.1 Science and the Scientific Method

Science is the process of *describing* the world around us. It is important to note that describing the world around us is not the same as *explaining* the world around us. Science aims to answer the question "How?" and not the question "Why?". As we develop our description of the physical world, you should remember this important distinction and resist the urge to ask "Why?".

The Scientific Method is a prescription for coming up with a description of the physical world that anyone can challenge and improve through performing experiments. If we come up with a description that can describe many observations, or the outcome of many different experiments, then we usually call that description a "Scientific Theory". We can get some insight into the Scientific Method through a simple example.

Imagine that we wish to describe how long it takes for a tennis ball to reach the ground after being released from a certain height. One way to proceed is to describe how long it takes for a tennis ball to drop 1 m, and then to describe how long it takes for a tennis ball to drop 2 m, etc. We could generate a giant table showing how long it takes a tennis ball to drop from any given height. Someone would then be able to perform an experiment to measure how long a tennis ball takes to drop from 1 m or 2 m and see if their measurement disagrees with the tabulated values. If we collected the descriptions for all possible heights, then we would effectively have a valid and testable scientific theory that describes how long it takes tennis balls to drop from any height.

Suppose that a budding scientist, let's call her Chloë, then came along and noticed that there is a pattern in the theory that can be described much more succinctly and generally than by using a giant table. In particular, suppose that she notices that, mathematically, the time, t, that it takes for a tennis ball to drop a height, h, is proportional to the square root of the height:

$$t \propto \sqrt{h}$$

Example 1-1

Use Chloë's Theory ($t \propto \sqrt{h}$) to determine how much longer it will take for an object to drop by $2\,\mathrm{m}$ than it would to drop by $1\,\mathrm{m}$.

Solution

When we have a proportionality law (with a $\propto$) sign, we can always change this to an equal sign by introducing a constant, which we will call k:

$$\begin{aligned} t &\propto \sqrt{h} \\ \rightarrow t &= k\sqrt{h} \end{aligned}$$

Let t_1 be the time to fall a distance $h_1 = 1\,\mathrm{m}$, and t_2 be the time to fall a distance $h_2 = 2\,\mathrm{m}$. In terms of our unknown constant, k, we have:

$$\begin{aligned} t_1 &= k\sqrt{h_1} = k\sqrt{(1\,\mathrm{m})} \\ t_2 &= k\sqrt{h_2} = k\sqrt{(2\,\mathrm{m})} \end{aligned}$$

By taking the ratio, $\frac{t_1}{t_2}$, our unknown constant k will cancel:

$$\begin{aligned} \frac{t_1}{t_2} &= \frac{\sqrt{(1\,\mathrm{m})}}{\sqrt{(2\,\mathrm{m})}} = \frac{1}{\sqrt{2}} \\ \therefore t_2 &= \sqrt{2}t_1 \end{aligned}$$

and we find that it will take $\sqrt{2} \sim 1.41$ times longer to drop by $2\,\mathrm{m}$ than it will by $1\,\mathrm{m}$.

Chloë's "Theory of Tennis Ball Drop Times" is appealing because it is succinct, and it also allows us to make **verifiable predictions**. That is, using this theory, we can predict that it will take a tennis ball $\sqrt{2}$ times longer to drop from $2\,\mathrm{m}$ than it will from $1\,\mathrm{m}$, and then perform an experiment to verify that prediction. If the experiment agrees with the prediction, then we conclude that Chloë's theory adequately describes the result of that particular experiment. If the experiment does not agree with the prediction, then we conclude that the theory is not an adequate description of that experiment, and we try to find a new theory.

Chloë's theory is also appealing because it can describe not only tennis balls, but the time it takes for other objects to fall as well. Scientists can then set out to continue testing her theory with a wide range of objects and drop heights to see if it describes those experiments as well. Inevitably, they will discover situations where Chloë's theory fails to adequately describe the time that it takes for objects to fall (can you think of an example?).

We would then develop a new "Theory of Falling Objects" that would include Chloë's theory that describes most objects falling, and additionally, a set of descriptions for the fall times for cases that are not described by Chloë's theory. Ideally, we would seek a new theory that would also describe the new phenomena not described by Chloë's theory in a succinct manner. There is of course no guarantee, ever, that such a theory would exist; it is just an optimistic hope of physicists to find the most general and succinct description of the physical world. This is a general difference between physics and many of the other sciences. In physics, one always tries to arrive at a succinct theory (e.g. an equation) that can describe many phenomena, whereas the other sciences are often very descriptive. For example, there is no succinct formula for how butterflies look; rather, there is a giant collection of observations of different butterflies.

This example highlights that applying the Scientific Method is an iterative process. Loosely, the prescription for applying the Scientific Method is:

1. Identify and describe a process that is not currently described by a theory.
2. Look at similar processes to see if they can be described in a similar way.
3. Improve the description to arrive at a "Theory" that can be generalized to make predictions.
4. Test predictions of the theory on new processes until a prediction fails.
5. Improve the theory.

Checkpoint 1-1

Fill in the blanks:

Physics is a branch of science that __________ the behaviour of the universe. When doing physics, we attempt to answer the question of __________ things work the way they do.

A) explains
B) describes
C) how
D) why

1.2 Theories, hypotheses and models

For the purpose of this textbook (and science in general), we introduce a distinction in what we mean by "theory", "hypothesis", and by "model". We will consider a "theory" to be a set of statements (or an equation) that gives us a broad description, applicable to several phenomena and that allows us to make verifiable predictions. For example, Chloë's Theory ($t \propto \sqrt{h}$) can be considered a theory. Specifically, we do not use the word theory in the context of "I have a theory about this..."

A "hypothesis" is a consequence of the theory that one can test. From Chloë's Theory, we have the hypothesis that an object will take $\sqrt{2}$ times longer to fall from 1 m than from 2 m. We can formulate the hypothesis based on the theory and then test that hypothesis. If the hypothesis is found to be invalidated by experiment, then either the theory is incorrect, or the hypothesis is not consistent with the theory.

A "model" is a situation-specific description of a phenomenon *based on a theory*, that allows us to make a specific prediction. Using the example from the previous section, our theory would be that the fall time of an object is proportional to the square root of the drop height, and a model would be applying that theory to describe a tennis ball falling by 4.2 m. From the model, we can form a testable hypothesis of how long it will take the tennis ball to fall that distance. It is important to note that a model will almost always be an approximation of the theory applied to describe a particular phenomenon. For example, if Chloë's Theory is only valid in vacuum, and we use it to model the time that it take for an object to fall at the surface of the Earth, we may find that our model disagrees with experiment. We would not necessarily conclude that the theory is invalidated, if our model did not adequately apply the theory to describe the phenomenon (e.g. by forgetting to include the effect of air drag).

This textbook will introduce the theories from Classical Physics, which were mostly established and tested between the seventeenth and nineteenth centuries. We will take it as given that readers of this textbook are not likely to perform experiments that challenge those well-established theories. The main challenge will be, given a theory, to define a model that describes a particular situation, and then to test that model. This introductory physics course is thus focused on thinking of "doing physics" as the task of correctly modelling a situation.

Emma's Thoughts

What's the difference between a model and a theory?
"Model" and "Theory" are sometimes used interchangeably among scientists. In physics, it is particularly important to distinguish between these two terms. A model provides an immediate understanding of something based on a theory.

For example, if you would like to model the launch of your toy rocket into space, you might run a computer simulation of the launch based on various theories of propulsion that you have learned. In this case, the model is the computer simulation, which describes what will happen to the rocket. This model depends on various theories that have been extensively tested such as Newton's Laws of motion, Fluid dynamics, etc.

- "Model": Your homemade rocket computer simulation
- "Theory": Newton's Laws of motion, Fluid dynamics

With this analogy, we can quickly see that the "model" and "theory" are not interchangeable. If they were, we would be saying that all of Newton's Laws of Motion depend on the success of your piddly toy rocket computer simulation!

Checkpoint 1-2

Models cannot be scientifically tested, only theories can be tested.

A) True
B) False

1.3 Fighting intuition

It is important to remember to fight one's intuition when applying the scientific method. Certain theories, such as Quantum Mechanics, are very counter-intuitive. For example, in Quantum Mechanics, an object can be described as being in two locations at the same time. In the Theory of Special Relativity, it is possible for two people to disagree on whether two events occurred at the same time. These particular prediction from these theories have not been invalidated by any experiment.

There is no requirement in science that a theory be "pretty" or intuitive. The only requirement is that a theory describe experimental data. One should then take care in not forcing one's preconceived notions into interpreting a theory. For example, Quantum Mechanics does not actually predict that objects can be in two locations at once, only that objects behave *as if* they were in two locations at once. A famous example is Schrödinger's cat, which can be modelled as being both alive and dead at the same time. However, just because we model it that way does not mean that it really is alive and dead at the same time.

1.4 The scope of Physics

Physics describes a wide range of phenomena within the physical sciences, ranging from the behaviour of microscopic particles that make up matter to the evolution of the entire Universe. We often distinguish between "classical" and "modern" physics depending on when the theories were developed, and we can further subdivide these areas of physics depending on the scale or the type of the phenomena that they describe.

The word physics comes from Ancient Greek and translates to "nature" or "knowledge of nature". The goal of physics is to develop theories from which mathematical models can be derived to describe our observations. One of the ambitious goals of physicists is to develop a single theory that describes all of nature, instead of having multiple theories to describe different categories of phenomena. This is in stark contrast to other fields of science, as

Rutherford famously quipped: "All science is either physics or stamp collecting". That is, physicists hope that there exists one single mathematical theory (like Chloë's theory of falling objects) that describes the entire physical world. In Biology, for example, this would not be a reasonable goal, as one needs to describe every single living being, and there is no overarching "theory of what all living things look like". Currently, physicists have been able to narrow down the number of theories required to describe all of the physical world to only three, which is impressive (the theory of gravity, the theory of the strong nuclear force, and physicists have now further unified the weak nuclear force with electromagnetism to make the "electroweak force").

1.4.1 Classical Physics

This textbook is focused on classical physics, which corresponds to the theories that were developed before 1905.

Mechanics

Mechanics describes most of our everyday experiences, such as how objects move, including how planets move under the influence of gravity. Isaac Newton was the first to formally develop a theory of mechanics, using his "Three Laws" to describe the behaviour of objects in our everyday experience. His famous work published in 1687, "Philosophiae Naturalis Principia Mathematica" ("The Principia") also included a theory of gravity that describes the motion of celestial objects.

Following the 1781 discovery of the planet Uranus by William Herschel, astronomers noticed that the orbit of the planet was not well described by Newton's theory. This led Urbain Le Verrier (in Paris) and John Couch Adams (in Cambridge) to predict the location of a new planet that was disturbing the orbit of Uranus rather than to claim that Newton's theory was incorrect. The planet Neptune was subsequently discovered by Le Verrier in 1846, one year after the prediction, and seen as a resounding confirmation of Newton's theory.

In 1859, Urbain Le Verrier also noted that Mercury's orbit around the Sun is different than that predicted by Newton's theory. Again, a new planet was proposed, "Vulcan", but that planet was never discovered and the deviation of Mercury's orbit from Newton's prediction remained unexplained until 1915, when Albert Einstein introduced a new, more complete, theory of gravity, called "General Relativity". This is a good example of the scientific method; although the discovery of Neptune was consistent with Newton's theory, it did not prove that the theory is correct, only that it correctly described the motion of Uranus. The discrepancy that arose when looking at Mercury ultimately showed that Newtons' theory of gravity fails to provide a proper description of planetary orbits in the proximity of very massive objects (Mercury is the closest planet to the Sun).

Checkpoint 1-3

What did the inability to find the planet Vulcan show:

A) It showed that Newton's model of Mercury was correct.
B) It showed that Newton's theory did not correctly describe the orbits of all planets.
C) It showed that the technology at the time was inadequate.
D) It showed that Einstein's theory of General Relativity was correct.

Electromagnetism

Electromagnetism describes electric charges and magnetism. At first, it was not realized that electricity and magnetism were connected. Charles Augustin de Coulomb published in 1784 the first description of how electric charges attract and repel each other. Magnetism was discovered in the ancient world, when people noticed that lodestone (rocks made from magnetized magnetite mineral) could attract iron tools. In 1819, Oersted discovered that moving electric charges could influence a compass needle, and several subsequent experiments were carried out to discover how magnets and moving electric charges interact.

In 1865, James Clerk Maxwell published "A Dynamical Theory of the Electromagnetic Field", wherein he first proposed a theory that unified electricity and magnetism as two facets of the same phenomenon. One important concept from Maxwell's theory is that light is an electromagnetic wave with a well-defined speed. This uncovered some potential issues with the theory as it required an absolute frame of reference in which to describe the propagation of light. Experiments in the late 1800s failed to detect the existence of this frame of reference.

1.4.2 Modern Physics

In 1905, Albert Einstein published three major papers that set the foundation for what we now call "Modern Physics". These papers covered the following areas that were not well-described by classical physics:

- A description of Brownian motion that implied that all matter is made of atoms.
- A description of the photoelectric effect that implied that light is made of particles.
- A description of the motion of very fast objects that implied that mass is equivalent to energy, and that time and distance are relative concepts.

In order to accommodate Einstein's descriptions, physicists had to dramatically re-formulate new theories.

Quantum mechanics and particle physics

Quantum mechanics is a theory that was developed in the 1920s to incorporate Einstein's conclusion that light is made of particles (or rather, quantized lumps of energy called quanta) and describe nature at the smallest scales. This could only be done at the expense of determinism, the idea that we can predict how particular situations evolve in time. This led to a theory that could only provide the *probabilities* that certain outcomes will be realized. Quantum mechanics was further refined during the twentieth century into Quantum Field Theory, which led to the Standard Model of particle physics that describes our current understanding of matter through the theories of the electroweak and strong forces.

The Special and General Theories of Relativity

In 1905, Einstein published his “Special Theory of Relativity”, which describes how light propagates at a constant speed without the need for an absolute frame of reference, thus solving the problem introduced by Maxwell. This required physicists to consider space and time on an equal footing (“space-time”), rather than two independent aspects of the natural world, and led to a flurry of odd, but verified, experimental predictions. One such prediction is that time flows slower for objects that are moving fast, which has been experimentally verified by flying precise atomic clocks on airplanes and satellites. In 1915, Einstein further refined his theory into General Relativity, which is our best current description of gravity and includes a description of Mercury’s orbit which was not described by Newton’s theory.

Checkpoint 1-4

Special relativity can be applied to which of these science fiction plots?

A) An eccentric duo travel back in time to alter the past.
B) An astronaut travelling near light speed for many years comes home to find that he has aged less than his family on Earth.
C) A superhero harnesses lightning to use as a weapon.

Cosmology and astrophysics

Cosmology describes processes at the largest scales and is mostly based on applying General Relativity to the scale of the Universe. For example, cosmology describes how our Universe started from the Big Bang and how large scale structures, such as galaxies and clusters of galaxies, have formed and evolved into our present day Universe.

Figure 1.1: A galaxy in the Coma cluster of galaxies (credit:NASA).

Astrophysics is focused on describing the formation and the evolution of stars, galaxies, and other “astrophysical objects” such as neutron stars and black holes.

Particle astrophysics

Particle astrophysics is a relatively new field that makes use of subatomic particles produced by astrophysical objects to learn both about the objects *and* about the particles. For example, the 2015 Nobel Prize in Physics was awarded to Art McDonald (a Canadian physicist from Queen’s University) for using neutrinos[1] produced by the Sun to both learn

[1] Neutrinos are the lightest subatomic particles that we know of

about the nature of neutrinos and about how the Sun works.

1.5 Thinking like a physicist

In a sense, physics can be thought of as the most fundamental of the sciences, as it describes the interactions of the smallest constituents of matter. In principle, if one can precisely describe how protons, neutrons, and electrons interact, then one can completely describe how a human brain thinks. In practice, the theories of particle physics lead to equations that are too difficult to solve for systems that include as many particles as a human brain. In fact, they are too difficult to solve exactly for even rather small systems of particles such as atoms bigger than helium (containing several protons, neutrons and electrons).

We have a number of other fields of science to cover complex systems of particles interacting. Chemistry can be used to describe what happens to systems consisting of many atoms and molecules. In a living being, it is too difficult to keep track of systems of atoms and molecules, so we use Biology to describe living systems.

One of the key qualities required to be an effective physicist is an ability to understand how to apply a theory and develop a model to describe a phenomenon. Just like any other skill, it takes practice to become good at developing models. Students that graduate with a physics degree are thus often sought for jobs that require critical thinking and the ability to develop quantitative models, which covers many fields from outside of physics such as finance or Big Data. This textbook thus tries to emphasize practice with developing models, while also providing a strong background in the theories of classical physics.

1.6 Summary

Key Takeaways

Science attempts to *describe* the physical world (it answers the question “How?”, not “Why?”).

The Scientific Method provides a prescription for arriving at theories that describe the physical world and that can be experimentally verified. The Scientific Method is necessarily an iterative process where theories are continuously updated as new experimental data are acquired. An experiment can only disprove a theory, not confirm it in any general sense.

Physics covers a wide scale of phenomena ranging from the Universe down to subatomic particles. Classical physics encompasses the theories developed before 1905, when Einstein introduced the need for Quantum Mechanics and the Theorie(s) of Relativity. One of the main goals of physics is to arrive at a single theory that describes all of our natural world. Currently, physicists require three theories to describe the natural world.

1.7 Thinking about the Material

Reflect and research

1. What particle helps to give mass to all of the massive elementary particles?
2. Name that physicist! Who was the first to propose that the universe is expanding?
3. Before discovering the CMBR (Cosmic Microwave Background Radiation), scientists Arno Penzias and Robert Wilson were trying to detect radio waves with very sensitive antennae. The very first time they heard a consistent, low noise on their detectors they discovered that it was (mostly) not the CMBR. What was causing most of this noise?
4. Physicist Lene Hau first slowed a beam of light to $17\,\mathrm{m/s}$ using a very cold, dilute gas of bosons. In 2001, how fast was she able to slow down the beam of light?
5. Think of two theories that you use in your every day life. (For example, when we wash our hands, we do so because of the germ theory of disease!)

1.8 Sample problems and solutions

1.8.1 Problems

Problem 1-1: Your friend Martin loves to explore "conspiracy theories". His favourite theory involves "Chem Trails". He tells you that the government is secretly using airliners to spread chemicals in the atmosphere for some unknown reason. (Solution)

a) Think of 2 ways in which you could objectively test Martin's theory.

b) After proposing your experiment to Martin, he claims that his theory cannot be invalidated by any experiment, no matter how scientifically rigorous the experiment is. Is Martin correct?

1.8.2 Solutions

Solution to problem 1-1:

a) You could do an investigation to see if the government is spreading chemicals, and try to find out why. You could make measurements of the contents in the atmosphere before and after an airline passes to see if any unexpected chemicals show up.

b) No he is not, as you just proposed two experiments that could invalidate his theory.

2 Gauss' Law

In this chapter, we take a detailed look at Gauss' Law applied in the context of the electric field. We have already encountered Gauss' Law briefly in Section **??** when we examined the gravitational field. Since the electric force is mathematically identical to the gravitational force, we can apply the same tools, including Gauss' Law, to model the electric field as we do the gravitational field. Many of the results from this chapter are thus equally applicable to the gravitational force.

Learning Objectives

- Understand the concept of flux for a vector field.
- Understand how to calculate the flux of a vector field through an open and a closed surface.
- Understand how to apply Gauss' Law quantitatively to determine an electric field.
- Understand how to apply Gauss' Law qualitatively to discuss charges on a conductor.

Think About It

A neutral spherical conducting shell encloses a point charge, Q, located at the centre of the shell. Due to separation of charge, the outer surface of the shell will acquire a net positive charge. What is the magnitude of that charge?

A) less than Q.
B) exactly Q.
C) more than Q.

2.1 Flux of the electric field.

Gauss' Law makes use of the concept of "flux". Flux is always defined based on:

- A surface.
- A vector field (e.g. the electric field).

and can be thought of as a measure of the number of field lines from the vector field that cross the given surface. For that reason, one usually refers to the "flux of the electric field through a surface". This is illustrated in Figure 2.1 for a uniform horizontal electric field, and a flat surface, whose normal vector, $\vec{A}$, is shown. If the surface is perpendicular to the field (left panel), and the field vector is thus parallel to the vector, $\vec{A}$, then the flux through that surface is maximal. If the surface is parallel to the field (right panel), then no field lines cross that surface, and the flux through that surface is zero. If the surface is rotated with respect to the electric field, as in the middle panel, then the flux through the surface is between zero and the maximal value.

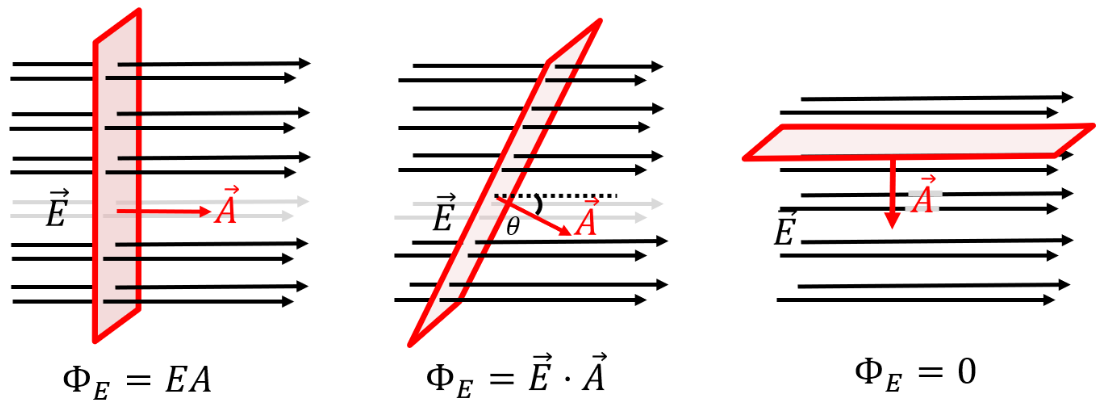

Figure 2.1: Flux of an electric field through a surface that makes different angles with respect to the electric field. In the leftmost panel, the surface is oriented such that the flux through it is maximal. In the rightmost panel, there are no field lines crossing the surface, so the flux through the surface is zero.

We define a vector, $\vec{A}$, associated with the surface such that the magnitude of $\vec{A}$ is equal to the area of the surface, and the direction of $\vec{A}$ is such that it is perpendicular to the surface, as illustrated in Figure 2.1. We define the flux, Φ_E, of the electric field, $\vec{E}$, through the surface represented by vector, $\vec{A}$, as:

$$\Phi_E = \vec{E} \cdot \vec{A} = EA\cos\theta$$

since this will have the same properties that we described above (e.g. no flux when $\vec{E}$ and $\vec{A}$ are perpendicular, flux proportional to number of field lines crossing the surface). Note that the flux is only defined up to an overall sign, as there are two possible choices for the direction of the vector $\vec{A}$, since it is only required to be perpendicular to the surface. By convention, we usually choose $\vec{A}$ so that the flux is positive.

Checkpoint 2-1

What are the units of electric flux?

A) Nm/C.
B) Vm.
C) V/m.
D) The units of flux depend on the dimensions of the charged object.

Example 2-1

A uniform electric field is given by: $\vec{E} = E\cos\theta\hat{x} + E\sin\theta\hat{y}$ throughout space. A rectangular surface is defined by the four points $(0,0,0)$, $(0,0,H)$, $(L,0,0)$, $(L,0,H)$. What is the flux of the electric field through the surface?

Solution

The surface that is defined corresponds to a rectangle in the xz plane with area $A = LH$. Since the rectangle lies in the xz plane, a vector perpendicular to the surface will be along the y direction. We choose the positive y direction, since this will give a positive number for the flux (as the electric field has a positive component in the y direction). The vector $\vec{A}$ is given by:

$$\vec{A} = A\hat{y} = LH\hat{y}$$

The flux through the surface is thus given by:

$$\begin{aligned}\Phi_E &= \vec{E} \cdot \vec{A} = (E\cos\theta\hat{x} + E\sin\theta\hat{y}) \cdot (LH\hat{y}) \\ &= ELH\sin\theta\end{aligned}$$

where one should note that the angle θ, in this case, is not the angle between $\vec{E}$ and $\vec{A}$, but rather the complement of that angle.

Discussion: In this example, we calculated the flux of a uniform electric field through a rectangle of area, $A = LH$. Since we knew the components of both the electric field vector, $\vec{E}$, and the surface vector, $\vec{A}$, we used their scalar product to determine the flux through the surface. In some cases, it is easier to work with the magnitude of the vectors and the angle between them to determine the scalar product (although note that in this example, the angle between $\vec{E}$ and $\vec{A}$ is $90^\circ - \theta$).

2.1.1 Non-uniform fields

So far, we have considered the flux of a uniform electric field, $\vec{E}$, through a surface, S, described by a vector, $\vec{A}$. In this case, the flux, Φ_E, is given by:

$$\Phi_E = \vec{E} \cdot \vec{A}$$

However, if the electric field is not constant in magnitude and/or in direction over the entire surface, then we divide the surface, S, into many infinitesimal surfaces, dS, and sum together (integrate) the fluxes from those infinitesimal surfaces:

$$\boxed{\Phi_E = \int \vec{E} \cdot d\vec{A}}$$

where, $d\vec{A}$, is the normal vector for the infinitesimal surface, dS. This is illustrated in Figure 2.2, which shows, in the left panel, a surface for which the electric field changes magnitude along the surface (as the field lines are closer in the lower left part of the surface), and, in the right panel, a scenario in which the direction (and magnitude) of the electric field vary along the surface.

In order to calculate the flux through the total surface, we first calculate the flux through an infinitesimal surface, dS, over which we assume that $\vec{E}$ is constant in magnitude and direction, and then, we sum (integrate) the fluxes from all of the infinitesimal surfaces together. Remember, the flux through a surface is related to the number of field lines that cross that surface; it thus makes sense to count the lines crossing an infinitesimal surface, dS, and then adding those together over all the infinitesimals surfaces to determine the flux through the total surface, S.

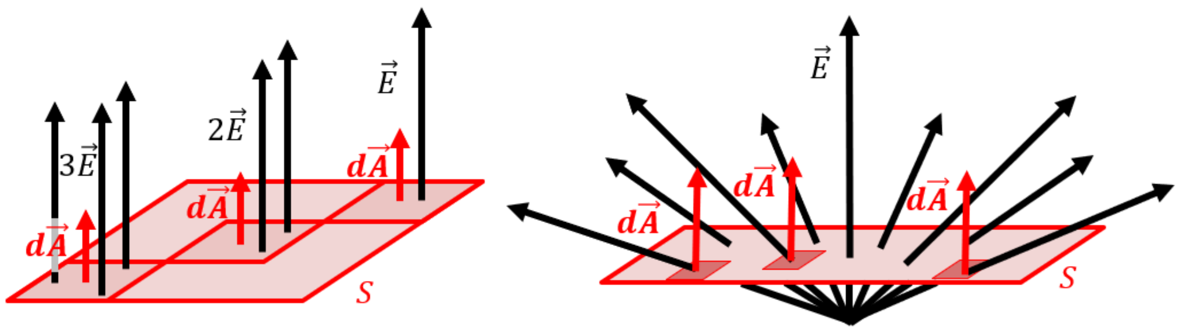

Figure 2.2: Examples of surfaces that need to be sub-divided in order to determine the net flux through them. The surface on the left must be subdivided because the electric field changes magnitude over the surface, whereas the one on the right needs to be subdivided because the angle between $\vec{E}$ and $d\vec{A}$ is not constant (and the magnitude of $\vec{E}$ also changes along the surface).

Example 2-2

An electric field points in the z direction everywhere in space. The magnitude of the electric field depends linearly on the x position in space, so that the electric field vector is given by: $\vec{E} = (a - bx)\hat{z}$, where, a, and, b, are constants. What is the flux of the electric field through a square of side, L, that is located in the xy plane?

Solution

We need to calculate the flux of the electric field through a square of side L in the xy plane. The electric field is always in the z direction, so the angle between $\vec{E}$ and $d\vec{A}$ (the normal vector for any infinitesimal area element) will remain constant.

We can calculate the flux through the square by dividing up the square into thin strips of length L in the y direction and infinitesimal width dx in the x direction, as illustrated in Figure 2.3. In this case, because the electric field does not change with y, the dimension of the infinitesimal area element in the y direction is finite (L). If the electric field varied both as a function of x and y, we would start with area elements that have infinitesimal dimensions in both the x and the y directions.

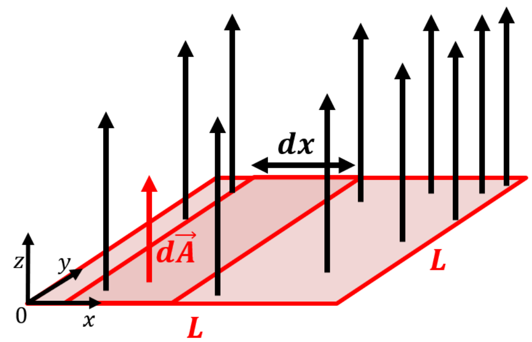

Figure 2.3: Dividing a square in the xy plane into thin strips of length L and width dx.

As illustrated in Figure 2.3, we first calculate the flux through a thin strip of area, $dA = Ldx$, located at position x along the x axis. Choosing, $d\vec{A}$, in the direction to give a positive flux, the flux through the strip that is illustrated is given by:

$$d\Phi_E = \vec{E} \cdot d\vec{A} = EdA = (ax - b)Ldx$$

where $\vec{E} \cdot d\vec{A} = EdA$, since the angle between $\vec{E}$ and $\vec{A}$ is zero. Summing together the fluxes from the strips, from $x = 0$ to $x = L$, the total flux is given by:

$$\Phi_E = \int d\Phi_E = \int_0^L (ax - b)Ldx = \frac{1}{2}aL^3 - bL^2$$

Discussion: In this example, we showed how to calculate the flux from an electric field that changes magnitude with position. We modelled a square of side, L, as being made of many thin strips of length, L, and width, dx. We then calculated the flux through each strip and added those together to obtain the total flux through the square.

2.1.2 Closed surfaces

One can distinguish between a "closed" surface and an "open" surface. A surface is closed if it completely defines a volume that could, for example, be filled with a liquid. A closed surface has a clear "inside" and an "outside". For example, the surface of a sphere, of a cube, or of a cylinder are all examples of closed surfaces. A plane, a triangle, and a disk are, on the other hand, examples of "open surfaces".

For a closed surface, one can unambiguously define the direction of the vector $\vec{A}$ (or $d\vec{A}$) as the direction that it is perpendicular to the surface and **points towards the outside**. Thus, the sign of the flux out of a closed surface is meaningful. The flux will be positive if there is a net number of field lines exiting the surface (since $\vec{E}$ and $\vec{A}$ will be parallel on average) and the flux will be negative if there is a net number of field lines entering the surface (as $\vec{E}$ and $\vec{A}$ will be anti-parallel on average). The flux through a closed surface is thus zero if the number of field lines that enter the surface is the same as the number of field lines that exit the surface.

When calculating the flux over a closed surface, we use a different integration symbol to show that the surface is closed:

$$\Phi_E = \oint \vec{E} \cdot d\vec{A}$$

which is the same integration symbol that we used for indicating a path integral when the initial and final points are the same (see for example Section **??**).

Checkpoint 2-2

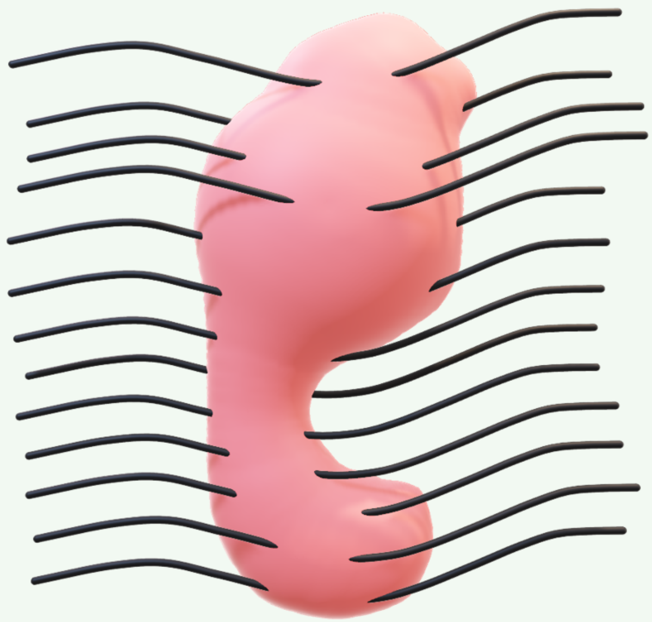

Figure 2.4: A non-uniform electric field flowing through an irregularly shaped object

A non-uniform electric field $\vec{E}$ flows through an irregularly-shaped two-dimensional object, as shown in figure 2.4. The flux through the irregularly-shaped object is...
A) positive.
B) zero.
C) negative.

Example 2-3

A negative electric charge, $-Q$, is located at the origin of a coordinate system. Calculate the flux of the electric field through a spherical surface of radius, R, that is centred at the origin.

Solution

Figure 2.5 shows the spherical surface of radius, R, centred on the origin where the charge $-Q$ is located.

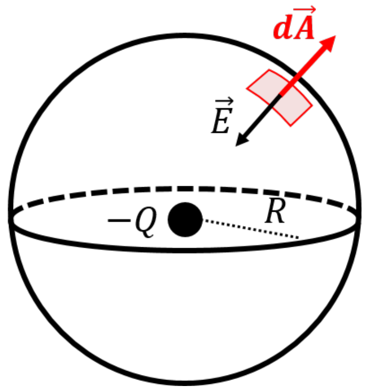

Figure 2.5: Calculating the flux through a spherical surface.

At all points along the surface, the electric field has the same magnitude:

$$E = \frac{1}{4\pi\epsilon_0}\frac{Q}{R^2}$$

as given by Coulomb's law for a point charge. Although the vector, $\vec{E}$, changes direction everywhere along the surface, it always makes the same angle ($-180°$) with the corresponding vector, $d\vec{A}$, at any particular location. Indeed, for a point charge, the electric field points in the radial direction (inwards for a negative charge) and is thus perpendicular to the spherical surface at all points. Since the surface is closed, the vector, $d\vec{A}$, points outwards anywhere on the surface. Thus, at any point on the surface, we can evaluate the flux through an infinitesimal area element, $d\vec{A}$:

$$d\Phi_E = \vec{E}\cdot d\vec{A} = EdA\cos(-180°) = -EdA$$

where the overall minus sign comes from the fact that, $\vec{E}$, and, $d\vec{A}$, are anti-parallel. The total flux through the spherical surface is obtained by summing together the fluxes through each area element:

$$\Phi_E = \oint d\Phi_E = \oint -EdA = -E\oint dA = -E(4\pi R^2)$$

where we factored, E, out of the integral, since the magnitude of the electric field is constant over the entire surface (a constant distance R from the charge). In the last equality, we recognized that, $\oint dA$, simply means "sum together all of the areas, dA, of the surface elements", which gives the total surface area of the sphere, $4\pi R^2$. The flux through the spherical surface is negative, because the charge is negative, and the field lines point towards $-Q$.

Using the value that we obtained for the magnitude of the electric field from Coulomb's Law, the total flux is given by:

$$\Phi_E = -E(4\pi R^2) = -\frac{1}{4\pi\epsilon_0}\frac{Q}{R^2}(4\pi R^2) = -\frac{Q}{\epsilon_0}$$

which, surprisingly, is independent of the radius of the spherical surface. Note that we used ϵ_0 instead of Coulomb's constant, k, since the result is cleaner without the extra factor of 4π.

Discussion: In this example, we calculated the flux of the electric field from a negative point charge through a spherical surface concentric with the charge. We found the flux to be negative, which makes sense, since the field lines go towards a negative charge, and there is thus a net number of field lines entering the spherical surface. Perhaps surprisingly, we found that the total flux through the surface does not depend on the radius of the surface! In fact, that statement is precisely Gauss' Law: the net flux out of a closed surface depends only on the amount of charge enclosed by that surface (and the constant, ϵ_0). Gauss' Law is of course more general, and applies to surfaces of any shape, as well as charges of any shape (whereas Coulomb's Law only holds for point charges).

2.2 Gauss' Law

Gauss' Law is a relation between the net flux through a closed surface and the amount of charge, Q^{enc}, in the volume enclosed by that surface:

$$\boxed{\oint \vec{E} \cdot d\vec{A} = \frac{Q^{enc}}{\epsilon_0}}$$

In particular, note that Gauss' Law holds true for **any** closed surface, and the shape of that surface is not specified in Gauss' Law. That is, we **can always choose the surface to use** when calculating the flux. For obvious reasons, we often call the surface that we choose a "gaussian surface". But again, this surface is simply a mathematical tool, there is no actual property that makes a surface "gaussian"; it simply means that we chose that surface in order to apply Gauss' Law. In Example 2-3 above, we confirmed that Gauss' Law is compatible with Coulomb's Law for the case of a point charge and a spherical gaussian surface.

Physically, Gauss' Law is a statement that field lines must begin or end on a charge (electric field lines original from positive charges and terminate on negative charges). Recall, flux is a measure of the net number of lines coming out of a surface. If there is a net number of lines coming out of a closed surface (a positive flux), that surface must enclose a positive charge from where those field lines originate. Similarly, if there are the same number of field lines entering a closed surface as there are lines exiting that surface (a flux of zero), then the surface encloses no charge. Gauss' Law simply states that the number of field lines exiting a closed surface is proportional to the amount of charge enclosed by that surface.

Primarily, Gauss' Law is a useful tool to determine the magnitude of the electric field from a given charge, or charge distribution. We usually have to use symmetry to determine the direction of the electric field vector. In general, the integral for the flux is difficult to evaluate, and Gauss' Law can only be used analytically in cases with a high degree of symmetry. Specifically, the integral for the flux is easiest to evaluate if:

1. **The electric field makes a constant angle with the surface**. When this is the

case, the scalar product can be written in terms of the cosine of the angle between $\vec{E}$ and $d\vec{A}$, which can be taken out of the integral if it is constant:

$$\oint \vec{E} \cdot d\vec{A} = \oint E \cos\theta dA = \cos\theta \oint EdA$$

Ideally, one has chosen a surface such that this angle is 0 or 180°.

2. **The electric field is constant in magnitude along the surface**. When this is the case, the integral can be simplified further by factoring out, E, and simply becomes an integral over dA (which corresponds to the total area of the surface, A):

$$\oint \vec{E} \cdot d\vec{A} = \cos\theta \oint EdA = E\cos\theta \oint dA = EA\cos\theta$$

Ultimately, the points above should dictate the choice of gaussian surface **so that** the integral for the flux is easy to evaluate. The choice of surface will depend on the symmetry of the problem. For a point (or spherical) charge, a spherical gaussian surface allows the flux to easily be calculated (Example 2-3). For a line of charge, as we will see, a cylindrical surface results is a good choice for the gaussian surface. Broadly, the steps for applying Gauss' Law to determine the electric field are as follows:

1. Make a diagram showing the charge distribution.
2. Use symmetry arguments to determine in which way the electric field vector points.
3. Choose a gaussian surface that goes through the point for which you want to know the electric field. Ideally, the surface is such that the electric field is constant in magnitude and always makes the same angle with the surface, so that the flux integral is straightforward to evaluate.
4. Calculate the flux, $\oint \vec{E} \cdot d\vec{A}$.
5. Calculate the amount of charge located within the volume enclosed by the surface, Q^{enc}.
6. Apply Gauss' Law, $\oint \vec{E} \cdot d\vec{A} = \frac{Q^{enc}}{\epsilon_0}$.

Example 2-4

An insulating sphere of radius, R, contains a total charge, Q, which is uniformly distributed through out its volume. Determine an expression for the electric field as a function of distance, r, from the centre of the sphere.

Solution

Note that this is identical, mathematically, as the derivation that is done in Section **??** for the case of gravity.

When applying Gauss' Law, we first need to think about symmetry in order to determine the direction of the electric field vector. We also need to think about all possible regions of space in which we need to determine the electric field. In particular, for this case, we need to determine the electric field both inside ($r \leq R$) and outside ($r \geq R$) of the charged sphere.

Figure 2.6 shows the charged sphere of radius R. If we consider the direction of the electric field outside the sphere (where $\vec{E}_{out}$ is drawn), we realize that it can only point in the radial direction (towards or away from the centre of the sphere), as this is the only choice that preserves the symmetry of the sphere. Being a sphere, the charge looks the same from all angles; thus, the electric field must also look the same from all angles, otherwise, there would be a preferred orientation for the sphere. The same argument holds for the electric field vector inside the sphere (drawn as $\vec{E}_{in}$).

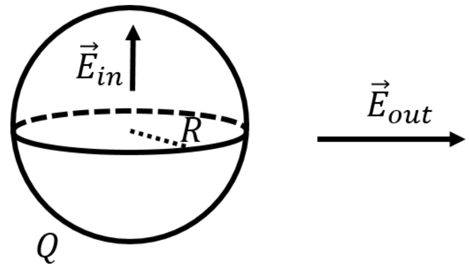

Figure 2.6: For a spherical charge distribution, the electric field inside and outside must point in the radial direction, by symmetry.

We now need to choose a gaussian surface that will make the flux integral easy to evaluate. Ideally, we can find a surface over which the electric field makes the same angle with the surface and over which the electric field is constant in magnitude. Again, based on the symmetry of the charge distribution, it is clear that a spherical surface of radius, r, will satisfy these properties.

We start by applying Gauss' Law outside the charge (with $r \geq R$) to determine the electric field, $\vec{E}_{out}$. Figure 2.7 shows our choice of spherical gaussian surface (labelled S) of radius, r, which is concentric with the spherical charge distribution of radius, R, and total charge, $+Q$.

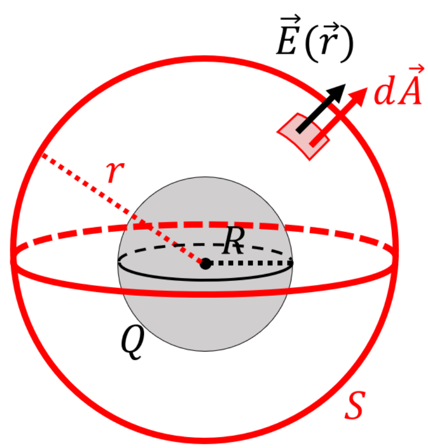

Figure 2.7: A spherical gaussian surface to determine the electric field outside a sphere of radius, R, holding charge, $+Q$.

In order to apply Gauss' Law, we need to calculate:

- the net flux through the surface.
- the charge in the volume enclosed by the surface.

The net flux through the surface is found identically as in Example 2-3, and is given by:

$$\Phi_E = \oint \vec{E} \cdot d\vec{A} = \oint E dA = E \oint dA = E(4\pi r^2)$$

where our choice of spherical surface led to $\vec{E} \cdot d\vec{A} = EdA$, since $\vec{E}$ and $d\vec{A}$ are always parallel. Furthermore, by symmetry, the electric field must be constant in magnitude along the whole surface, or the spherical symmetry would be broken. This allowed us to factor the E out of the integral, leaving us with, $\oint dA$, which is simply the area of our gaussian spherical surface, $4\pi r^2$.

The gaussian surface with $r \geq R$ encloses the whole charged sphere, so the charge enclosed is simply the charge of the sphere, $Q^{inc} = Q$. Applying Gauss' Law allows us to determine the magnitude of the electric field:

$$\begin{aligned}
\oint \vec{E} \cdot d\vec{A} &= \frac{Q^{enc}}{\epsilon_0} \\
E(4\pi r^2) &= \frac{Q^{enc}}{\epsilon_0} \\
\therefore E &= \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}
\end{aligned}$$

which is the same as the electric field a distance r from a point charge. Thus, from the outside, a spherical charge distribution leads to the same electric field as if the charge were concentrated at the centre of the sphere.

Next, we determine the magnitude of the electric field inside the charged sphere. In this case, we choose a spherical gaussian surface of radius $r \leq R$, that is concentric with the sphere, as illustrated by the surface labelled, S, that is shown in Figure 2.8.

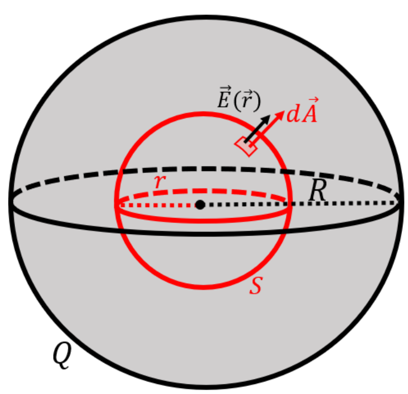

Figure 2.8: A spherical gaussian surface to determine the electric field outside a sphere of radius, R, holding charge, +Q.

The flux integral is trivial again, since the electric field always makes the same angle with the gaussian surface, and the magnitude of the electric field is constant in magnitude along the surface:

$$\Phi_E = \oint \vec{E} \cdot d\vec{A} = \oint E dA = E \oint dA = E(4\pi r^2)$$

In this case, however, the charge in the volume enclosed by the gaussian surface is less than Q, since the whole charge is not enclosed. We are told that the charge is distributed uniformly throughout the spherical volume of radius R. We can thus define a volume charge density, ρ, (charge per unit volume) for the sphere:

$$\rho = \frac{Q}{V} = \frac{Q}{\frac{4}{3}\pi R^3}$$

The volume enclosed by the gaussian surface is $\frac{4}{3}\pi r^3$, thus, the charge, Q^{enc}, contained in that volume is given by:

$$Q^{enc} = \frac{4}{3}\pi r^3 \rho = \frac{4}{3}\pi r^3 \frac{Q}{\frac{4}{3}\pi R^3} = Q\frac{r^3}{R^3}$$

Finally, we apply Gauss' Law to find the magnitude of the electric field inside the sphere:

$$\begin{aligned}
\oint \vec{E} \cdot d\vec{A} &= \frac{Q^{enc}}{\epsilon_0} \\
E(4\pi r^2) &= \frac{Q^{enc}}{\epsilon_0} = \frac{Q}{\epsilon_0}\frac{r^3}{R^3} \\
\therefore E &= \frac{Q}{4\pi\epsilon_0 R^3} r
\end{aligned}$$

Note that the electric field increases linearly with radius inside of the charge sphere, and then decreases with radius squared outside of the sphere. Also, note that at the centre of the sphere, the electric field has a magnitude of zero, as expected from symmetry.

Discussion: In this example, we showed how to use Gauss' Law to determine the electric field inside and outside of a uniformly charged sphere. We recognized the spherical symmetry of the charge distribution and chose to use a spherical surface in order to apply Gauss' Law. This, in turn, allowed the flux to be easily calculated. We found that outside the sphere, the electric field decreases in magnitude with radius squared, just as if the entire charge were concentrated at the centre of the sphere. Inside the sphere, we found that the electric field is zero at the centre, and increases linearly with radius.

Checkpoint 2-3

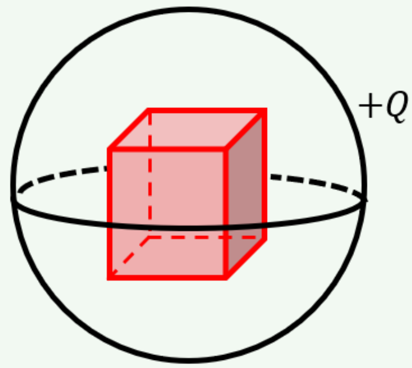

Figure 2.9: A charged spherical shell with a cubic device inside of it.

A thin charged spherical shell of radius r has a uniformly distributed charge of $+Q$. Inside of the hollow inside of the shell is a cubic device which measures flux as shown in Figure 2.9. What is the total flux through the cubic device?

A) $\frac{Q}{12\pi}$ Vm
B) $\frac{Q}{2\pi}$ Vm
C)
D) $0\,\text{Vm}$.

Example 2-5

An infinitely long straight wire carries a uniform charge per unit length, λ. What is the electric field at a distance, R, from the wire?

Solution

We start by making a diagram of the charge distribution, as in Figure 2.10, so that we can use symmetry arguments to determine the direction of the electric field vector. At any point in space, the electric field vector must be radial (point to/from the centre of the wire) and in the plane perpendicular to the wire. If this were not the case, one would be able to look at the electric field to determine a preferred direction (either around the wire, if the field were not radial, or upwards/downwards, if the field were not perpendicular to the wire).

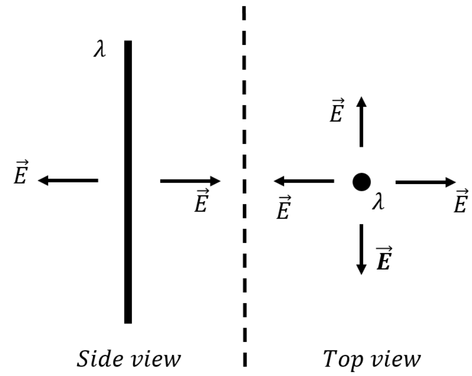

Figure 2.10: An infinite line of charge carrying uniform charge per unit length, λ. The left panel shows a side view and the right panel a view from above. The electric field must be in the radial direction or there would be a preferred direction.

Next, we need to choose a gaussian surface in order to apply Gauss' Law. A convenient choice is a cylinder (a "pill box") of radius, R, and length, L, as shown in Figure 2.11, as this goes through a point that is a distance, R, from the wire (where we are asked for the electric field). At all points on the cylindrical surface, the electric field vector is either perpendicular or parallel to the surface.

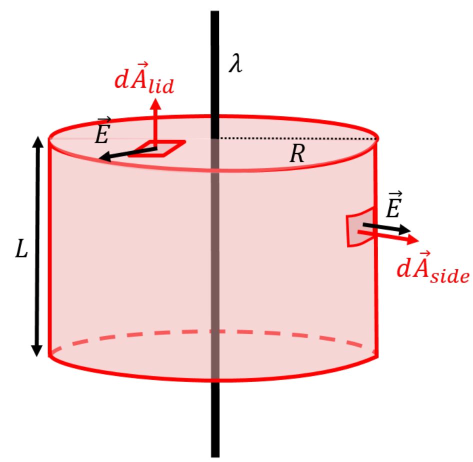

Figure 2.11: A cylindrical gaussian surface is used to calculate the flux from an infinite line of charge.

We can think of the cylindrical surface as being composed of three surfaces: 2 disks on either end (the lids of the pill box), and the curved surface that makes up the side of the cylinder. The flux through the entire cylindrical surface will be the sum of the fluxes through the two lids plus the flux through the side:

$$\oint \vec{E}\cdot d\vec{A} = \int_{side} \vec{E}\cdot d\vec{A} + \int_{lid} \vec{E}\cdot d\vec{A} + \int_{lid} \vec{E}\cdot d\vec{A}$$

where you should note that the closed integral ($\oint$) was separated into three normal integrals ($\int$) corresponding to the three "open" surfaces that make up the closed surface. Again, remember that the flux is proportional to the net number of field lines exiting/entering the closed surface, so it make sense to count those lines over the three open surfaces and add them together to get the total number for the closed surface.

The flux through the lids is identically zero, since the electric field is perpendicular to $d\vec{A}$ everywhere on the lids. The total flux is thus equal to the flux through the curved side surface, for which the electric field vector is always parallel to $d\vec{A}$, and for which the electric field vector is constant in magnitude:

$$\oint \vec{E}\cdot d\vec{A} = \int_{side}\vec{E}\cdot d\vec{A} = \int_{side} EdA = E\int_{side} dA = E(2\pi RL)$$

where we recognized that the side surface can be unfolded into a rectangle of height, L, and width, $2\pi R$, corresponding to the circumference of the cylinder, so that the area is given by $A = 2\pi RL$.

Next, we determine the charge inside the volume enclosed by the surface. Since the cylinder encloses a length, L, of wire, the enclosed charge is given by:

$$Q^{enc} = \lambda L$$

where λ is the charge per unit length on the wire. Putting this altogether into Gauss' Law gives us the electric field at a distance, R, from the wire:

$$\begin{aligned}\oint \vec{E}\cdot d\vec{A} &= \frac{Q^{enc}}{\epsilon_0}\\ E(2\pi RL) &= \frac{\lambda L}{\epsilon_0}\\ \therefore E &= \frac{\lambda}{2\pi\epsilon_0 R}\end{aligned}$$

Note that this is the same result that we obtained in Example **??** when we took the limit of the finite line of charge having infinite length.

Discussion: In this example, we applied Gauss' Law to determine the electric field at a distance from an infinitely long charged wire. We used symmetry to argue that the field should be radial and in the plane perpendicular to the wire, and recognized that a cylindrical gaussian surface would exploit the symmetry so that the flux can easily be calculated. We obtained the same result as we did from integrating Coulomb's Law in Example **??**. However, using Gauss' Law was much less work than integrating Coulomb's Law.

Checkpoint 2-4

Why can't Gauss' Law be applied to a finite wire?

A) Gauss' Law can be applied to a finite wire.

B) Because the flux of a finite wire is undefined.

C) Because we do not know the charge density of a finite wire.

D) Because the symetry argument does not hold.

Josh's Thoughts

Gauss' Law allows us to choose a gaussian surface, but which surface should we choose? This is a difficult question to answer, because it is highly circumstancial. Generally, it is useful to choose a gaussian surface which requires the least complex math to solve, and it is often advised to use the symmetry of the charge to avoid taking an integral. If symmetry can be exploited such that $\vec{E}$ has a constant magnitude at every $d\vec{A}$ of the gaussian surface, then $\int \vec{E} \cdot d\vec{A}$ will be equal to EA. This is why gaussian surfaces are often of the same form as the charged object they are enclosing.

For example: if I were to enclose a cylindrical continuous charge, it would be reasonable to enclose the continuously charged cylinder with a cylindrical gaussian surface, as shown in Figure 2.12

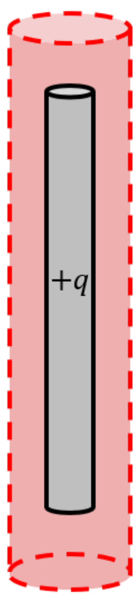

Figure 2.12: .

When dealing with point charges which have no shape, it is often the best choice to choose a spherical gaussian surface because the electric field is emitted isotropically from the point charge, as shown in Figure 2.13

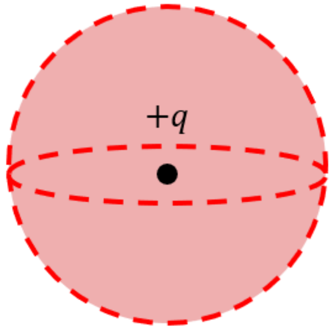

Figure 2.13: .

Finally, there are some cases of less useful choices for gaussian surfaces. While not wrong, they may produce an equation which requires unnecessary integration. These cases will still provide a correct answer if the situation is modelled correctly, but there is often a more simple method of obtaining the answer. Suppose I enclosed a spherical charge with a long cylinder, as shown in Figure 2.14. The electric field will be stronger near the middle of the cylinder's length than at the centre of its endcaps, which means that $\vec{E}$ is not constant in $\int \vec{E} \cdot d\vec{A}$, so the integral cannot be simplified to EA. A better choice for a gaussian surface in this case would be a sphere, which exploits the symmetry of the charge distribution and provides an $\vec{E}$ of constant magnitude for all $d\vec{A}$. Figures 2.3 and 2.2 give other examples of when we cannot assume Φ to be equal to EA

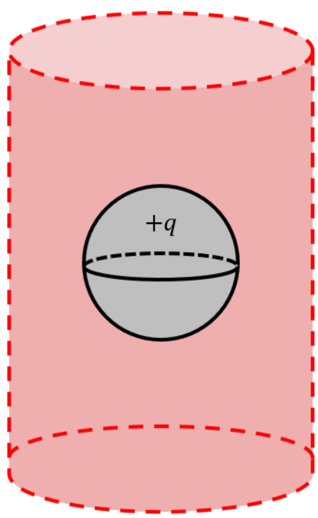

Figure 2.14: .

Example 2-6

Determine the electric field above an infinitely large plane of charge with uniform surface charge per unit area, σ.

Solution

Figure 2.15 shows a portion of the infinite plane. The electric field vector must be perpendicular to the plane or a preferred direction could otherwise be inferred from the direction of the electric field. We can also argue that the horizontal components of the electric field will cancel everywhere above the plane, since the plane is infinite. The electric field will point away from (towards) the plane, if the charge is positive (negative).

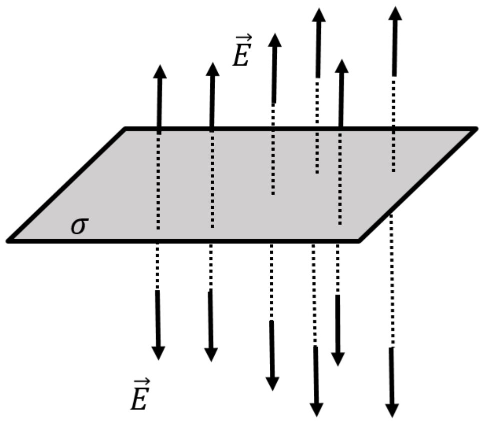

Figure 2.15: The electric field above an infinite plane with uniform charge per unit area, σ, must be perpendicular to the plane.

A cylindrical or box-shaped gaussian surface would both lead to the flux integral being easy to calculate, as illustrated in Figure 2.16. Indeed, since the electric field is perpendicular to the plane, only the parts of the surface that are parallel to the plane (the lids on the cylinder, the two horizontal planes in the box) will have a net flux through them.

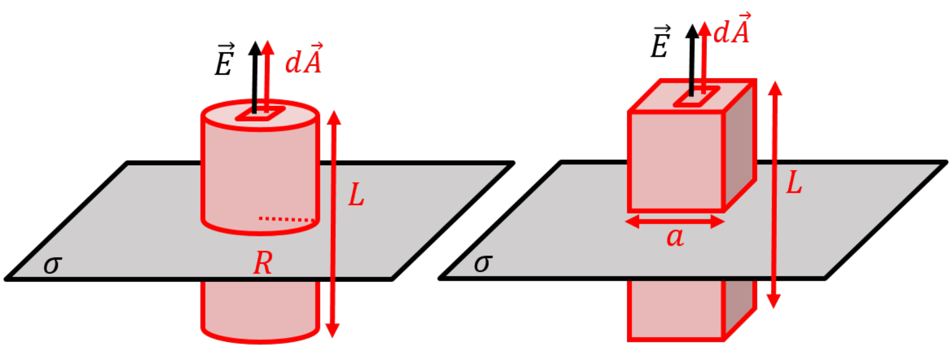

Figure 2.16: A cylindrical surface or a box are both good choices for a gaussian surface above a plane, since only the parts of the surface parallel to the plane will have net flux through them.

Let us choose a box (right panel of Figure 2.16) of length, L, with a square cross-section of side, a. We place the box such that the plane intersects the centre of the box (although this is not required, since we already know that the electric field will not depend on distance from the plane). The flux through the box is simply the flux through the two horizontal planes (of area a^2):

$$\oint \vec{E} \cdot d\vec{A} = \int_{top} EdA + \int_{bottom} EdA = 2Ea^2$$

The box encloses a section of the plane with area a^2, so that the net charge enclosed by the surface is:

$$Q^{enc} = \sigma a^2$$

Applying Gauss' Law allows us to determine the magnitude of the electric field:

$$\begin{aligned}
\oint \vec{E} \cdot d\vec{A} &= \frac{Q^{enc}}{\epsilon_0} \\
2Ea^2 &= \frac{\sigma a^2}{\epsilon_0} \\
\therefore E &= \frac{\sigma}{2\epsilon_0}
\end{aligned}$$

which is the same result that we found in Example **??**.

Discussion: In this example, we used Gauss' Law to determine the electric field above an infinite plane. We found that we had a choice of gaussian surfaces (cylinder, box) that allowed us to apply Gauss' Law. We found the same result that we had found in Example **??** where we had integrated Coulomb's Law (twice, once for a ring of charge, then for a disk, then took the limit of the disk radius going to infinity). Again, we see that in configurations with a high degree symmetry, Gauss' Law can be very straightforward to apply.

2.3 Charges in a conductor

We can use Gauss' Law to understand how charges arrange themselves on a conductor. Consider (again) an infinite plane that carries a total charge per unit area, σ, similar to what we considered in Example 2-6. In this case, we explicitly consider the plane to be a conductor and to have a finite thickness. If we zoom into the plane, we will see that the charges will migrate to the surface of the plane, as illustrated in Figure 2.17, where the plane is seen edge on. Thus, the **charge density at the surface is half of the total charge density** of the plane.

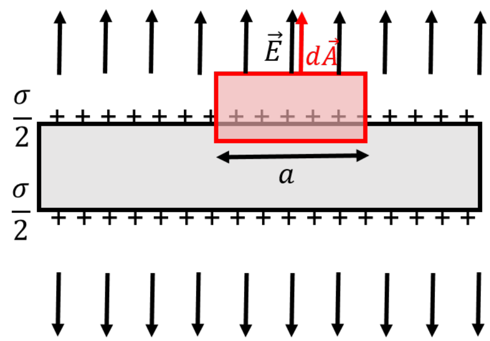

Figure 2.17: Cross-section of a conducting plane where the charges migrate to the surface. A box-shaped gaussian surface is also shown as seen from the side (the third dimension of the box is perpendicular to the plane of the page).

To determine the electric field near the plane, we choose a gaussian surface that is a box (as in Example 2-6), but require the lower end of the box to go through the plane, as illustrated in Figure 2-6. With this choice of gaussian surface, only the top surface (area a^2) will have flux through it, since the **electric field inside a conductor must be zero**[1]. The total flux is given by:

$$\oint \vec{E}\cdot d\vec{A} = \int_{top} EdA = Ea^2$$

The charge enclosed is given by:

$$Q^{enc} = \frac{\sigma}{2}a^2$$

where we used the fact that only half of the charges are inside the volume enclosed by our gaussian surface, so that the charge per unit area is half ($\frac{\sigma}{2}$) of that for the entire plane. Applying Gauss' Law, we find that the electric field is given by:

$$\begin{aligned}
\oint \vec{E}\cdot d\vec{A} &= \frac{Q^{enc}}{\epsilon_0}\\
Ea^2 &= \frac{\sigma a^2}{2\epsilon_0}\\
\therefore E &= \frac{\sigma}{2\epsilon_0} \quad \text{(Field above an infinite plane)}
\end{aligned}$$

as before, but the factor of 2 now came from the charge density, rather than from the fact that two of the faces of the box had non-zero flux (as was the case in Example 2-6). We can generalize this result to determine the electric field near the surface of any conductor. Very close to the surface of any object, one can consider the surface as being similar to an infinite plane. If that surface carries charge per unit area, σ, then the electric field just above the surface is given by:

$$E = \frac{\sigma}{\epsilon_0} \quad \text{(Field near a conducting surface)}$$

[1]Since charges can freely move in a conductor, they will move until there is no reason to move. Eventually, the charges accumulate in such a way that the net field in the conductor is zero. For a plane, this means that half of the charges will move to each side, as illustrated.

In this case, there is no factor of 2, because the charge density in this equation is the charge density at the surface of the conductor. In the previous equation, the charge density on the surface of the conducing plane was $\frac{\sigma}{2}$.

Consider, now, a neutral spherical conducting shell, as shown from the side in the left panel of Figure 2.18. When a charge, $+Q$, is placed at the centre of the shell (right panel), charges inside the shell will move until the field in the shell is identically zero. The negative charges will move towards the inner surface (as they are attracted to $+Q$) and positive charges will be repelled onto the outer surface, under the influence of the electric field created by $+Q$ (shown in the diagram as $\vec{E}_Q$). Eventually, the separation of charges will lead to an electric field (shown in the diagram as $\vec{E}_\sigma$) in the opposite direction. The charges will stop moving once the total electric field in the conductor is zero (when the two fields cancel exactly everywhere in the conductor).

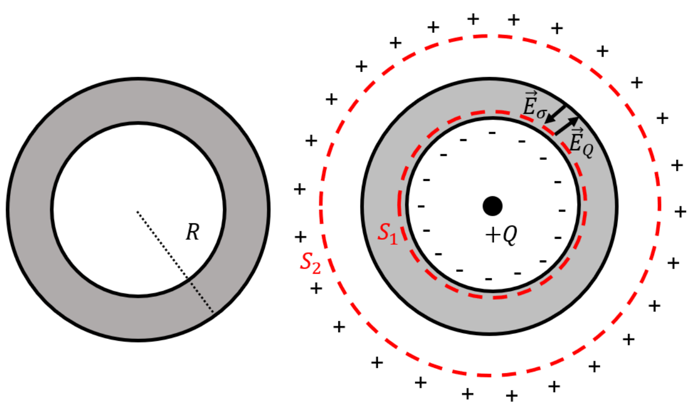

Figure 2.18: Left: a neutral conducting spherical shell (seen edge on). Right: A positive charge, $+Q$, placed at the centre of the shell. Charges in the shell will separate in order to keep the electric field inside the conductor zero.

We can use Gauss' Law to determine the amount of charge that has accumulated on the inner surface. Consider the gaussian spherical surface, S_1, in Figure 2.18, that is concentric with the shell and has a radius such that the surface is just inside the shell. Since the electric field is zero inside the shell, the flux out of the gaussian surface must be zero. By Gauss' Law, the amount of charge enclosed by the surface must also be zero. Thus, a total charge, $-Q$, will have accumulated on the inner surface of the conductor (since $Q^{enc} = -Q + Q = 0$). Because one cannot just create charge from nothing, there must be an equal amount of opposite charge, $+Q$, on the outer surface of the shell. This is true of any conducting material with a cavity inside of it: if you place a charge $+Q$ in the cavity, a charge, $-Q$ will accumulated on the inner surface and a charge, $+Q$, will accumulate on the outer surface.

If we now consider the flux out of the surface, S_2, outside of the shell, the net charge enclosed will be $Q^{enc} = +Q - Q + Q = +Q$. The flux out of the spherical surface of radius, say, r, is then given by:

$$\oint \vec{E} \cdot d\vec{A} = E(4\pi r^2)$$

and the electric field, from Gauss' Law, is simply that of a point charge, $+Q$:

$$E = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}$$

and the shell has no effect on the field in regions where there is no conducting material from the shell. Right at the surface of the shell (outer radius, R), the surface charge density is given by:

$$\sigma = \frac{Q}{4\pi r^2}$$

Above, we found the electric field at the surface of a conductor that carries charge per unit area, σ, to be:

$$E = \frac{\sigma}{\epsilon_0}$$

which is clearly the same result that we obtained using the spherical surface, S_2:

$$E = \frac{\sigma}{\epsilon_0} = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}$$

Note that we found the electric field using Gauss' Law only in this last case, and found it to be equal to the electric field that one obtains from Coulomb's law. Thus, Gauss' Law only works if the field has an "inverse square law" dependence. If Gauss' Law does not provide the correct electric field, then the force does not depend on $1/r^2$. Gauss' Law can be used to make extremely stringent tests of whether the force goes as $1/r^2$ or deviates from this model.

2.4 Interpretation of Gauss' Law and vector calculus

In this section, we provide a little more theoretical background and intuition on Gauss' Law, as well as its connection to vector calculus. Very generally, Gauss' Law is a statement that connects a property of a vector field to the "source" of that field. We think of mass as the source for the gravitational field, and we think of charge as the source for the electric field. The property of the field that we considered in this case was its "flux out of a closed surface".

Recall that determining the flux of a field out of a closed surface is equivalent to counting the net number of field lines that exit that closed surface. Field lines must start on a positive charge and must end on a negative charge. Thus, if there is a net number of field lines exiting the surface, there must be a positive charge in the volume defined by the surface (a "source" of field lines). If there is a net number of field lines entering the surface, then the volume defined by the surface must enclose a negative charge (a "sink" of field lines). Gauss' Law is simply a statement that the number of field lines entering/exiting a closed surface is proportional to the amount of charge enclosed in that volume.

The flux out of a closed surface is tightly connected to the vector calculus concept of "divergence", which describes whether field lines are diverging (spreading out or getting closer together). When a point charge is present, field lines will emanate radially from that point charge; in other words, they will diverge. We say that the electric field has non-zero divergence if there is a source of the electric field in that position of space. The key difference between the concept of divergence and that of "flux out of a closed surface", is that divergence is a local property of the field (it is true at a point), whereas the flux out of a surface must be calculated using a finite volume and makes it challenging to define the field at a specific position. Gauss's Law defined using flux is thus not as useful for describing how the field changes at specific positions, and is usually limited to situations with a high degree of symmetry.

The divergence, $\nabla \cdot \vec{E}$, of a vector field, $\vec{E}$, at some position is defined as:

$$\nabla \cdot \vec{E} = \frac{\partial E}{\partial x} + \frac{\partial E}{\partial y} + \frac{\partial E}{\partial z}$$

and corresponds to the sum of three partial derivatives evaluated at that position in space. Gauss' Theorem (also called the Divergence Theorem) states that:

$$\int_V \nabla \cdot \vec{E} = \oint_S \vec{E} \cdot d\vec{A}$$

where the V (S) on the integral indicate whether the sum (integral) should be carried out over a volume, V, or over a closed surface, S, as we have practised in this chapter. While it is not important at this level to understand the theorem in detail, the point is that one can convert a "flux over a closed surface" into an integral of the divergence of the field. In other words, we can convert a global property (flux) to a local property (divergence). Gauss' Law in terms of divergence can be written as:

$$\boxed{\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0}} \quad \text{(Local version of Gauss' Law)}$$

where ρ is the charge per unit volume at a specific position in space. This is the version of Gauss' Law that is usually seen in advanced textbooks and in Maxwell's unified theory of electromagnetism. This version of Gauss's Law relates a local property of the field (its divergence) to a local property of charge at that position in space (the charge per unit volume at that position in space). If we integrate both sides of the equation over volume, we recover the original formulation of Gauss' Law: the left hand side, by the Divergence Theorem, leads to flux when integrated over volume, whereas on the right hand side, the integral over volume of charge per unit volume, ρ, will give the total charge enclosed in that volume, Q^{enc}:

$$\int_V \left(\nabla \cdot \vec{E}\right) dV = \int_V \left(\frac{\rho}{\epsilon_0}\right) dV$$
$$\oint_S \vec{E} \cdot d\vec{A} = \frac{Q^{enc}}{\epsilon_0}$$

2.5 Summary

Key Takeaways

We can define the **flux** of a uniform and constant vector field, $\vec{E}$, through a flat surface, as:

$$\Phi_E = \vec{E} \cdot \vec{A} = EA\cos\theta$$

where, $\vec{A}$, is a vector that is perpendicular to the surface with a magnitude equal to the area of that surface, and, θ, is the angle between, $\vec{A}$ and $\vec{E}$. The flux of a field through a surface is proportional to the number of field lines that cross that surface. If the surface is parallel to the field ($\vec{A}$ and $\vec{E}$ are thus perpendicular), the flux through that surface is zero (no field lines cross the surface, the scalar product is zero).

If $\vec{E}$ and $\vec{A}$ change over the surface ($\vec{E}$ and/or $\vec{A}$ change magnitude and/or direction along the surface), then we treat the surface as being made of infinitesimal surface elements over which the two vectors are constant. We define a vector $d\vec{A}$ to be perpendicular to the surface element with an infinitesimal area, dA. The total flux is then obtained by summing the fluxes through each surface element:

$$\Phi_E = \int \vec{E} \cdot d\vec{A} = \int EdA\cos\theta$$

Note that the direction of the vector $d\vec{A}$ (or $\vec{A}$) is ambiguous, as one can choose either of two directions perpendicular to a surface. Usually, one chooses the direction of $\vec{A}$ so that the flux is positive (i.e. $\vec{A}$ has a component parallel to $\vec{E}$). However, if the surface is "closed" (that is, it defines a volume), then we always choose the direction of $d\vec{A}$ so that it points outwards from the surface (since the surface encloses a volume, one can define an "inside" and an "outside").

In the case of the electric field, Gauss' Law relates the flux of the electric field from a closed surface to the amount of charge, Q^{enc}, contained in the volume enclosed by that surface:

$$\oint \vec{E} \cdot d\vec{A} = \frac{Q^{enc}}{\epsilon_0}$$

Physically, Gauss' Law is a statement that field lines must begin or end on a charge (electric field lines original from positive charges and terminate on negative charges). If there is a net number of lines coming out of a closed surface (a positive flux), that surface must enclose a positive charge from where those field lines originate. Similarly, if there are the same number of field lines entering a closed surface as there are lines exiting that surface (a flux of zero), then the surface encloses no charge. Gauss' Law states that the number of field lines exiting a closed surface is proportional to the amount of charge enclosed by that surface.

Gauss' Law is useful to determine the electric field. However, this can only be done analytically for charge distributions with a very high degree of symmetry. This is because the flux integral is not usually easy to evaluate unless:

1. **The electric field makes a constant angle with the surface**. When this is the case, the scalar product can be written in terms of the cosine of the angle between $\vec{E}$ and $d\vec{A}$, which can be taken out of the integral if it is constant:

$$\oint \vec{E} \cdot d\vec{A} = \oint E \cos\theta dA = \cos\theta \oint E dA$$

2. **The electric field is constant in magnitude along the surface**. When this is the case, the integral can be simplified further by factor out E, and simply becomes an integral over dA (which corresponds to the total area of the surface, A):

$$\oint \vec{E} \cdot d\vec{A} = \cos\theta \oint E dA = E\cos\theta \oint dA = EA\cos\theta$$

Note that Gauss' Law does not specify a closed surface over which to calculate the flux; it holds for any surface. We can thus choose a surface that will make the flux integral easy to evaluate - we call this choice a "gaussian surface" (not because it has some special property, but because we chose that surface to apply Gauss' Law). A procedure for applying Gauss' Law to determine the electric field at some point in space can be written as:

1. Make a diagram showing the charge distribution.
2. Use symmetry arguments to determine in which way the electric field vector points.
3. Choose a gaussian surface that goes through the point for which you want to know the electric field. Ideally, the surface is such that the electric field is constant in magnitude and always makes the same angle with the surface, so that the flux integral is straightforward to evaluate.
4. Calculate the flux, $\oint \vec{E} \cdot d\vec{A}$.
5. Calculate the amount of charge in the volume enclosed by the surface, Q^{enc}.
6. Apply Gauss' Law, $\oint \vec{E} \cdot d\vec{A} = \frac{Q^{enc}}{\epsilon_0}$.

We showed how Gauss' Law can be used to understand and quantify how charges arrange themselves on a conductor, in such a way that the electric field is zero everywhere in the conductor. Finally, we briefly introduced a more modern version of Gauss' Law that uses divergence instead of flux:

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0}$$

This last version has the advantage that it relates a local property of the field (divergence) to a local property of charge (charge density at some position in space).

Important Equations

Gauss' Law:

$$\Phi = \frac{Q_{enc}}{\epsilon_0}$$

$$\Phi = \int \vec{E} \cdot d\vec{A}$$

Important Definitions

Electric flux: A measurement of an electric field through a closed surface. SI units: [Vm]. Common variable(s): Φ.

Vacuum permittivity: A fundamental physical constant which states the dielectric permittivity of a vacuum. SI units: [Fm^{-1}]. Common variable(s): ϵ_0.

2.6 Thinking about the material

Reflect and research

1. Could Gauss' law be applied to magnetism? Why or why not?
2. What are Maxwell's equations?
3. How are measurements of flux used in environmental research?

To try at home

1.

To try in the lab

1. Propose an experiment to measure the charge of an object using Gauss' law.
2. Propose an experiment to measure the electric field of a charged object, then compare your experimental results to the theoretical results predicted calculated by Gauss' law.
3. Simulate the surface charge distribution on the inside and outside of a conducting cubic shell which encloses a point charge.

2.7 Sample problems and solutions

2.7.1 Problems

Problem 2-1: Consider a sphere which has a charge density of $\rho = ar^2$ and has a radius R. What is the electric field at the distances $0 \leq d \leq r$ and $r < d$ from the centre of the sphere? (Solution)

Problem 2-2: Consider two conducting plates. Both plates have a hollow circle of radius r at their centre. One plate is a square on the outside and the other is a triangle on the outside, both of the outside shapes have a side length of L. A point charge of charge $+Q$ is placed at the centre of the hollowed out circle of both plates. (Solution)

a) What is the electric field outside of the shells?
b) What is the average linear charge density on the inner and outter surfaces of the shells?
c) Which sections of the two plates would have the largest charge density?

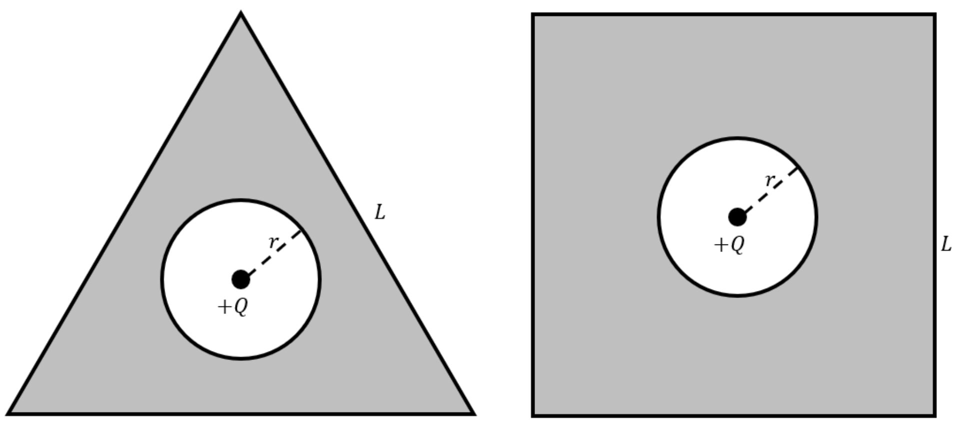

Figure 2.19: A triangular and square shell, both with a hollowed out circular centre and a point charge.

2.7.2 Solutions

Solution to problem 2-1:

First, we must find the total charge of the sphere. To do this, we will need to integrate ρ over the volume of the sphere:

$$Q = \int_0^R \rho dV$$
$$Q = \int_0^R 4\pi r^4 dr$$
$$Q = \frac{4}{5}\pi R^5$$

Now that we have our charge, Q, we can find the enclosed charge at any distance r by taking the ratio of the volumes, then multiplying it by the total charge of the sphere:

$$Q_{enc} = \frac{Q\frac{4}{3}\pi r^3}{\frac{4}{3}\pi R^3}$$
$$Q_{enc} = \frac{Qr^3}{R^3}$$
$$Q_{enc} = \frac{4\pi r^3 R^2}{5}$$

Now that we have our enclosed charge for any distance r, we must apply Gauss' law:

$$\Phi = EA = \frac{Q_{enc}}{\epsilon_0}$$
$$E(4\pi r^2) = \frac{4\pi r^3 R^2}{5\epsilon_0}$$
$$E = \frac{rR^2}{5\epsilon_0}$$

Which gives our answer for the magnitude of the electric field felt at a distance r from the centre of the shell when $0 \leq r \leq R$.

To find the electric field at a distance r from the centre of the sphere when $R < r$, we will apply the same technique, but will instead set the enclosed charge to be constant at $r = R$:

$$\Phi = EA = \frac{Q_{enc}}{\epsilon_0}$$
$$E(4\pi r^2) = \frac{4\pi R^5}{5\epsilon_0}$$
$$E = \frac{\pi R^5}{5\epsilon_0 r^2}$$

Which gives our final answer.

Solution to problem 2-2:

(a) The conducting shells have no net charge, so the only charge in the system is the point charge Q, which means that the electric field outside of the shells is simply $E = \frac{kQ}{r^2}$.

(b) Let's begin with the shell that has a triangle on the outside. We will use Gauss' law to determine the charge density of the inner and outer shells. To do this, we will draw a circle within the shell, S_1 and a triangle outside of the outer shell, S_2:

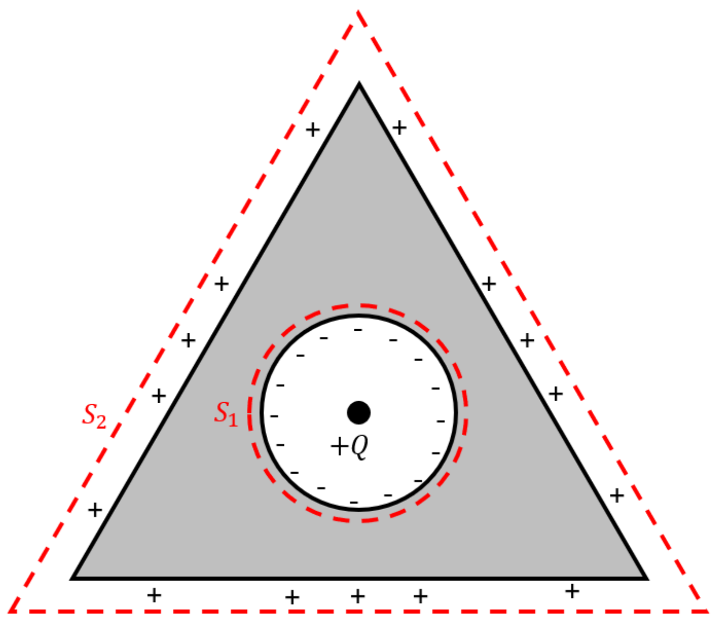

Figure 2.20: A solution to the triangular conducting shell

When considering S_1 know that the electric field inside of the shell is 0, so we also know that the flux will be zero. This means that the point charge on the inside of the shell will be equal and opposite to the sum of the surface charges on the inner shell. From here, we divide the net charge by the circumference of the inner shell to determine the linear charge density:

$$\lambda_{circle} = \frac{-Q}{2\pi r}$$

When considering S_2, we know that the $Q_{enc} = +Q$, which means that the total linear charge on the outer triangle will be $+Q$ such that it cancels the $-Q$ along the inner circle, leaving the point charge, $+Q$. The sum of charges would be $Q_{enc} = Q_{point} + Q_{triangle} - Q_{circle}$. Knowing this, we must divide the total charge on the outter shell by the sum of the length of each of the triangle's sides in order to find the average linear charge density:

$$\lambda_{circle} = \frac{Q}{3L}$$

Now, we must solve for the inner and outer linear charge densities of the conducting shell with a square outter surface. We will begin this process by drawing a circle within the shell, S_1, and a square outside of the shell, S_2

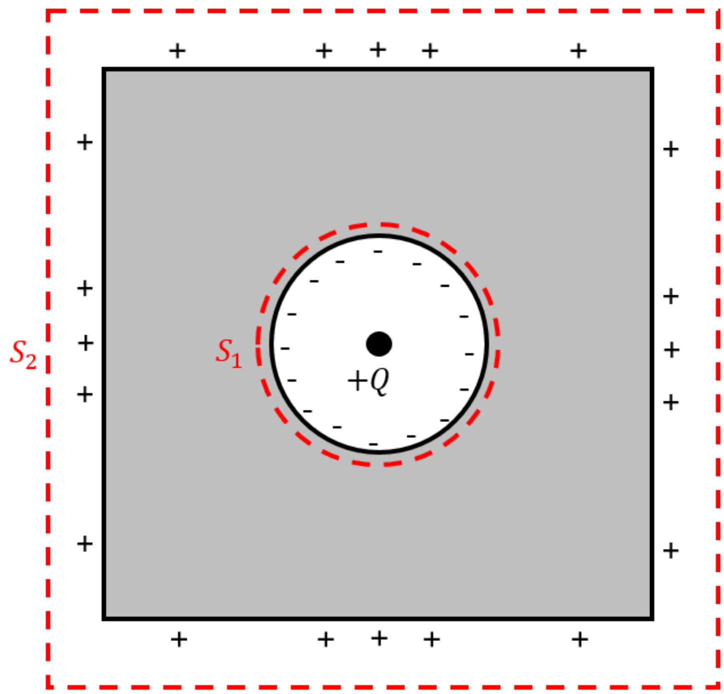

Figure 2.21: A solution to the square conducting shell

For S_1, the circle is treated as it was while solving the trianglular shell. The electric field is also 0 within the square conducting shell, so we know that the average linear charge density is $\frac{-Q}{2\pi r}$.
When considering S_2, we know that Q_{enc} is $+Q$, so we know that the total charge on the square surface of the shell will be $+Q$. This leave us with the following average linear charge density:

$$\lambda_{square} = \frac{Q}{4L}$$

(c) These plates are charged by the electric field generated by the point charge held within them, which means that the linear charge density of the two plates will be highest at the points along the outer sides which are the shortest distance from the point charge. These points occur in the triangular and square plates at the point which is the greatest distance from their vertices, as shown in Figure 2.22.

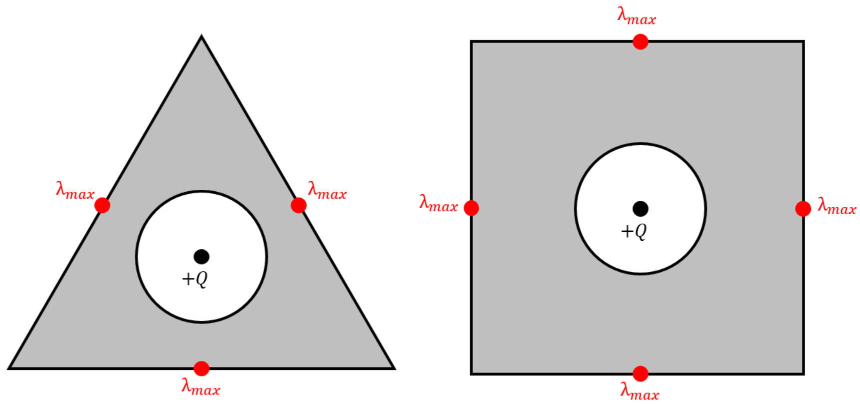

Figure 2.22: A solution to the square conducting shell

3 Electric potential

In this chapter, we develop the concept of electric potential energy and electric potential. This will allow us to describe the motion of charges using energy instead of forces. We will also introduce the capacitor, a common circuit component that is used to store charge.

Learning Objectives

- Understand the difference between electrical potential energy and electric potential.
- Understand how to calculate stored electrostatic potential energy.
- Understand how to calculate the electric potential difference between two points near a point charge or a distribution of charges.
- Understand how to use electric potential to determine electrical potential energy.
- Understand how to determine electric potential from electric field.
- Understand how to determine electric field from electric potential.
- Understand how to model a capacitor.

Think About It

A proton and an electron are both accelerated by the 110 V electric potential difference from your outlet. Which particle has the highest speed?

A) The proton.
B) The electron.
C) They will have the same speed, since they were accelerated by the same potential difference.

3.1 Electric potential energy

Review Topics

- Section **??** on conservative forces.
- Section **??** on the derivation of gravitational potential energy.

Mathematically, Coulomb's Law for the electric force is identical to Newton's Universal Theory of Gravity for the gravitational force. The electric force is thus conservative, and the work done by the electric force on a charge, q, when the charge moves from position, A, in space to some other position, B, cannot depend on the path taken. Since the work done by the electric force only depends on the location of the initial (A) and final (B) positions, we can define an electrical potential energy function, $U(\vec{r})$, that depends on position. The work done by the electric force, $\vec{F}^E$, on a charge in going from position, A (defined by position vector $\vec{r}_A$), to position, B (defined by position vector $\vec{r}_B$), can be written as:

$$W = \int_A^B \vec{F}^E \cdot d\vec{r} = -\Delta U = -\left[U(\vec{r}_B) - U(\vec{r}_A)\right] \tag{3.1}$$

In order to determine the function, $U(\vec{r})$, we can choose a path over which the integral for work is easy to calculate. Consider the work done by the electric force from a point charge, $+Q$, exerted on a charge, $+q$, when $+q$ moves from a distance r_A to a distance r_B from the centre of $+Q$, as illustrated in Figure 3.1.

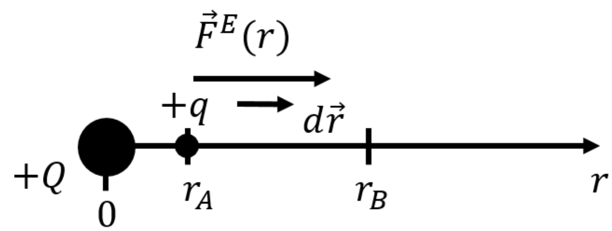

Figure 3.1: Calculating the work done on a charge $+q$ by the electric force exerted by charge $+Q$ when charge $+q$ moves from a distance r_A to a distance r_B from the centre of charge $+Q$.

Placing $+Q$ at the origin of a coordinate system, the force exerted on charge, $+q$, when it is located at position, $\vec{r}$, is given by:

$$\vec{F}^E = k\frac{Qq}{r^2}\hat{r}$$

The work done by the electric force when $+q$ moves from A to B is given by:

$$\begin{aligned} W &= \int_A^B \vec{F}^E \cdot d\vec{r} = \int_{\vec{r}_A}^{\vec{r}_B} \left(k\frac{Qq}{r^2}\hat{r}\right) \cdot d\vec{r} = kQq\int_{r_A}^{r_B} \frac{1}{r^2}dr \\ &= kQq\left[\frac{-1}{r}\right]_{r_A}^{r_B} = -\left(\frac{kQq}{r_B} - \frac{kQq}{r_A}\right) \end{aligned}$$

where we noted that since $\vec{F}^E$ and $d\vec{r}$ are parallel, their scalar product is simply the product of their magnitudes. By comparing with Equation 3.1, we can identify the potential energy, $U(\vec{r})$, of a charge, $+q$, located at a relative position, $\vec{r}$, from a point charge, $+Q$, as:

$$\boxed{U(\vec{r}) = \frac{kQq}{r} + C}$$

where the potential energy is only defined up to some constant, C, which cancels when we take the difference in potential energy between two positions. Note that this is very similar to the function for the gravitational potential energy of a mass, m, a distance, r, from a mass, M (see Section **??**).

The potential energy function that we derived above remains the same if one or both of the charges change sign, as the derivation did not depend on the sign of the charges, q and Q, as changing the sign of one charge changes the direction of the force. For example, a positive charge, $+q$, near a negative charge, $-Q$, would have negative electric potential energy with the choice $C = 0$, in exact analogy with gravity.

3.1.1 Electrostatic potential energy

When we hold two, say, positive charges together a distance, r, apart, we need to exert a force on the charges in order to keep the charges in place (as they repel each other). If we release the charges, they will move apart from each other, and eventually all of the stored electric potential energy is converted into kinetic energy. The energy that was originally stored in this "system" of two charges is called "electrostatic potential energy". In this section, we show how to model the energy stored in a collection of point charges.

Consider a single positive charge, q_1, located at the origin of empty space. Since there are no other charges present, it does not "cost" us any energy to place that charge there - we do not need to do any work. If we now bring in a second positive charge, q_2, and place it a distance, r_{12}, from q_1 (Figure 3.2), we will need to do work since q_1 exerts a force on q_2. If we define zero potential energy to be at infinity (choosing $C = 0$ for electric potential energy), the work, W_{q2}, that we must do on q_2 to bring it from infinity to a distance, r_{12}, from q_1 is given by the corresponding change in potential energy of q_2:

$$W_{q2} = \Delta U = U_{final} - U_{initial} = k\frac{q_1 q_2}{r_{12}} - 0 = k\frac{q_1 q_2}{r_{12}}$$

Note that the work is done by us (not by the electric field), so it has the same sign as the change in potential energy (we must do positive work to increase potential energy). The work that we did corresponds to the same amount of electrostatic potential energy stored in this arrangement of two charges (the only source of that stored electrostatic potential energy is the work that we did on the charge q_2).

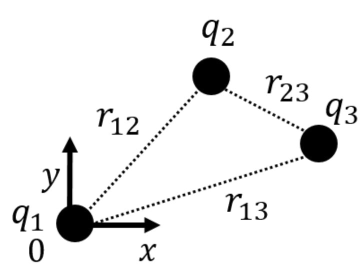

Figure 3.2: Three positive charges arranged together will store a certain amount of electrostatic potential energy.

Now, we bring in a third positive charge, q_3, also from infinitely far away, as illustrated in Figure 3.2. In order to bring in q_3, we need to do work against the forces exerted by both q_1 and q_2. Suppose that we place q_3 a distance r_{13} from q_1 and r_{23} from q_2. Then, the amount of work done by us to bring in q_3 is given by:

$$W_{q3} = k\frac{q_1 q_3}{r_{13}} + k\frac{q_2 q_3}{r_{23}}$$

and the total electrostatic energy stored in the system of three charges is given by the sum of the work done to place q_2 and the work done to place q_1:

$$E = W_{q1} + W_{q2} + W_{q3} = 0 + k\frac{q_1 q_2}{r_{12}} + k\frac{q_1 q_3}{r_{13}} + k\frac{q_2 q_3}{r_{23}}$$

If we have any number of charges (positive and negative), we can always calculate the stored electrostatic energy by proceeding in a similar fashion.

Checkpoint 3-1

Four charges of varying magnitude are fixed in position. If the electric potential energy stored in the system were to be calculated, how many terms would be in the sum?

A) Four
B) Two
C) One
D) Six

3.2 Electric potential

As you recall, we defined the **electric field**, $\vec{E}(\vec{r})$, to be the **electric force per unit charge**. By defining an electric field everywhere in space, we were able to easily determine the force on any test charge, q, whether the test charge is positive or negative (since the sign of q will change the direction of the force vector, $q\vec{E}$):

$$\begin{aligned}\vec{E}(\vec{r}) &= \frac{\vec{F}^E(\vec{r})}{q}\\ \therefore \vec{F}^E(\vec{r}) &= q\vec{E}(\vec{r})\end{aligned}$$

Similarly, we define the **electric potential**, $V(\vec{r})$, to be the **electric potential energy per unit charge**. This allows us to define electric potential, $V(\vec{r})$, everywhere in space, and then determine the potential energy of a specific charge, q, by simply multiplying q with the electric potential at that position in space.

$$\begin{aligned}V(\vec{r}) &= \frac{U(\vec{r})}{q}\\ \therefore U(\vec{r}) &= qV(\vec{r})\end{aligned}$$

The S.I. unit for electric potential is the "volt", (V). Electric potential, $V(\vec{r})$, is a scalar field whose value is "the electric potential" at that position in space. A positive charge, $q = 1\,\text{C}$, will thus have a potential energy of $U = 10\,\text{J}$ if it is located at a position in space where the electric potential is $V = 10\,\text{V}$, since $U = qV$. Similarly, a negative charge, $q = -1\,\text{C}$, will have negative potential energy, $U = -10\,\text{J}$ at the same location.

Since only differences in potential energy are physically meaningful (as change in potential energy is related to work), **only changes in electrical potential are physically meaningful** (as electric potential is related to electric potential energy). A difference in electric potential is commonly called a "voltage". One often makes a clear choice of where the electric potential is zero (typically the ground, or infinitely far away), so that the term voltage is used to describe potential, V, instead of difference in potential, ΔV; this should only be done when it is clear where the location of zero electric potential is defined.

We can describe a free-falling mass by stating that the mass moves from a region where it has high gravitational potential energy to a region of lower gravitational potential energy under the influence of the force of gravity (the force associated with a potential energy always acts in the direction to decreases potential energy). The same is true for electrical potential energy: **charges will always experience a force in a direction to decrease their electrical potential energy**. However, positive charges will experience a force driving them from regions of high electric potential to regions of low electric potential, whereas negative charges will experience a force driving them from regions of low electric potential to regions of higher electric potential. This is because, for negative charges, the change in potential energy associated with moving through space, ΔU, will be the negative of the corresponding change in electric potential, $\Delta U = q\Delta V$, since the charge, q, is negative.

Checkpoint 3-2

Electric potential increases along the x-axis. A proton and an electron are placed at rest at the origin; in which direction do the charges move?

A) the proton moves towards lower electric potential, while the electron moves towards higher electric potential.
B) the proton moves towards higher electric potential, while the electron moves towards lower electric potential.
C) the proton and electron move towards lower electric potential.
D) the proton and electron move towards higher electric potential.

If the only force exerted on a particle is the electric force, and the particle moves in space such that the electric potential changes by ΔV, we can use conservation of energy to determine the corresponding change in kinetic energy of the particle:

$$\begin{aligned}
\Delta E &= \Delta U + \Delta K = 0 \\
\Delta U &= q\Delta V \\
&\boxed{\therefore \Delta K = -q\Delta V}
\end{aligned}$$

where ΔE is the change in total mechanical energy of the particle, which is zero when energy is conserved. The kinetic energy of a positive particle increases if the particle moves from a region of high potential to a region of low potential (as ΔV would be negative and q is positive), and vice versa for a negative particle. This makes sense, since a positive and negative particle feel forces in opposite directions.

In order to describe the energies of particles such as electrons, it is convenient to use a different unit of energy than the Joule, so that the quantities involved are not orders of magnitude smaller than 1. A common choice is the "electron volt", eV. One electron volt corresponds to the energy acquired by a particle with a charge of e (the charge of the electron) when it is accelerated by a potential difference of 1 V:

$$\begin{aligned}
\Delta E &= q\Delta V \\
1\,\text{eV} &= (e)(1\,\text{V}) = 1.6\times 10^{-19}\,\text{J}
\end{aligned}$$

An electron that has accelerated from rest across a region with a 150 V potential difference across it will have a kinetic of 150 eV = 2.4×10^{-17} J. As you can see, it is easier to describe the energy of an electron in electron volts than Joules.

Checkpoint 3-3

A particle moves from an electric potential of −260 V to an electric potential of −600 V and loses kinetic energy. What is the charge of this particle?

A) Neutral.
B) It could have a positive or a negative charge.
C) Positive.
D) Negative.

Josh's Thoughts

It is often useful in physics to take previously learned concepts and compare them to new ones, in this case, gravitational potential energy and electric potential energy can be compared to help understand the physical meaning of electric potential.

Suppose an object with a large mass, M, is sitting in space. Now place an object of a much smaller mass, m at any distance r from the centre of M. The gravitational potential energy of the small mass is dependant on r, m, and M, and could be calculated with the following formula:

$$U_g = \frac{GMm}{r}$$

Which is very similar to the formula for electrical potential energy:

$$U(\vec{r}) = \frac{kQq}{r}$$

Now, if we were to remove the mass m from its position, we would no longer have an object with gravitational potential energy. However, we could still describe the gravitational potential for the point r, which would produce gravitational potential energy when given any mass m. This is the gravitational equivalent to electric potential, and could be calculated with the following formula:

$$V_g = \frac{U_g}{m}$$

Which is also very similar to the formula for electric potential:

$$V_E = \frac{U_E}{q}$$

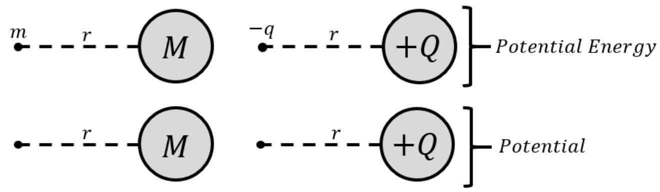

Figure 3.3: Gravitational potential energy and gravitational potential (left) next to its electrical analog (right)

Example 3-1

A proton and an electron move from a region of space where the electric potential is 20 V to a region of space where the electric potential is 10 V. If the electric force is the only force exerted on the particles, what can you say about their change in speed?

Solution

The two particles move from a region of space where the electric potential is 20 V to a region of space where the electric potential is 10 V. The change in electric potential experienced by the particles is thus:

$$\Delta V = V_{final} - V_{initial} = (10\,\text{V}) - (20\,\text{V}) = -10\,\text{V}$$

and we take the opportunity to emphasize that one should be very careful with signs when using potential. The change in potential energy of the proton, with charge $q = +e$, is thus:

$$\Delta U_p = q\Delta V = (+e)(-10\,\text{V}) = -10\,\text{eV}$$

The potential energy of the proton thus decreases by 10 eV (which you can easily convert to Joules). Since we are told that no other force is exerted on the particle, the total mechanical energy of the particle (kinetic plus potential energies) must be constant. Thus, if the potential energy decreased, then the kinetic energy of the proton has increased by the same amount, and **the proton's speed increases**.

The change in potential energy of the electron, with charge $q = -e$, is thus:

$$\Delta U_e = q\Delta V = (-e)(-10\,\text{V}) = 10\,\text{eV}$$

The potential energy of the electron thus increases by 10 eV. Again, the mechanical energy of the electron is conserved, so that an increase in potential energy results in the same decrease in kinetic energy and **the electron's speed decreases**.

Discussion: By using the electric potential, V, we modelled the change in electric potential energy of a proton and an electron as they both moved from one region of space to another.

We found that when a **proton moves from a region of high electric potential to a region of lower electric potential, its potential energy decreases**. This is because the proton has a positive charge and a decrease in electric potential will also result in a decrease in potential energy. Since no other forces are exerted on the proton, the proton's kinetic energy must increase. Because the potential energy of the proton decreases, the proton is moving in the same direction as the electric force, and the electric force does positive work on the proton to increase its kinetic energy.

Conversely, we found that when an **electron moves from a region of high electric potential to a region of lower electric potential, its potential energy increases**. This is because it has a negative charge and a decrease in electrical potential thus results in an increase in potential energy. Since no other forces are exerted on the electron, the electron's kinetic energy must decrease, and the electron slows down. This makes sense, since the force that is exerted on an electron will be in the opposite direction from the force exerted on a proton.

3.2.1 Electric potential from electric field

At the beginning of Section 3.1, we determined the potential energy of a point charge, q, in the presence of another point charge, Q (Figure 3.1). This was done by calculating the work done by the Coulomb (electric) force exerted by charge Q on q. We can write the same integral for the work done by the electric force on q, but using the electric field, $\vec{E}$, to write the force:

$$W = \int_A^B \vec{F}^E \cdot d\vec{r} = \int_A^B q\vec{E} \cdot d\vec{r} = q\int_A^B \vec{E} \cdot d\vec{r}$$

where we recognized that the charge, q, is constant and can come out of the integral. The integral that is left is thus the work done by the electric field, $\vec{E}$, *per unit charge*. In other words, this is the negative change in electric potential:

$$W = q\int_A^B \vec{E} \cdot d\vec{r} = -q\Delta V = -q\left[V(\vec{r}_B) - V(\vec{r}_A)\right]$$

$$\boxed{\therefore \Delta V = V(\vec{r}_B) - V(\vec{r}_A) = -\int_A^B \vec{E} \cdot d\vec{r}}$$

which allows us to easily determine the change in electric potential associated with an electric field. Note that this result is general and does not require the electric field to be that of a point charge, and can be used to determine the electric potential associated with any electric field. We can also specify a function for the potential, up to an arbitrary constant,

C, (think definite versus indefinite integrals):

$$V(\vec{r}) = -\int \vec{E} \cdot d\vec{r} + C$$

The relation between electric potential and electric field is analogous to the relation between electric potential energy and electric force:

$$\Delta V = V(\vec{r}_B) - V(\vec{r}_A) = -\int_A^B \vec{E} \cdot d\vec{r}$$
$$\Delta U = U(\vec{r}_B) - U(\vec{r}_A) = -\int_A^B \vec{F}^E \cdot d\vec{r}$$

as the bottom equation is just q times the first equation. We can think of electric potential being to potential energy what electric field is to electric force. Electric potential and electric field are electric potential energy and electric force, *per unit charge*, respectively.

For a point charge, Q, located at the origin, the electric field at some position, $\vec{r}$, is given by Coulomb's Law:

$$\vec{E} = \frac{kQ}{r^2}\hat{r}$$

The potential difference between location A (at position $\vec{r}_A$) and location B (at position $\vec{r}_B$), as in Figure 3.1, is given by:

$$\Delta V = -\int_A^B \vec{E} \cdot d\vec{r} = -\int_{\vec{r}_A}^{\vec{r}_B} \frac{kQ}{r^2}\hat{r} \cdot d\vec{r} = -\left(\frac{kQ}{r_B} - \frac{kQ}{r_A}\right)$$

and we note that we can write a function for the electric potential, $V(\vec{r})$, at a distance r from a point charge, Q, as:

$$V(\vec{r}) = \frac{kQ}{r} + C$$

where C is an arbitrary constant. This, of course, is identical to the result that we obtained earlier, for the potential energy of a charge, q, a distance, r, from Q.

$$U(\vec{r}) = qV(\vec{r}) = \frac{kQq}{r} + C'$$

where the constant, $C' = qC$, does not have any physical impact. Often, as is the case for gravity, one chooses the constant $C = 0$. This choice corresponds to defining potential energy to be zero at infinity. Equivalently, this corresponds to choosing infinity to be at an electric potential of $0\,\text{V}$.

Checkpoint 3-4

What causes a positively charged particle to gain speed when it is accelerated through a potential difference?:

A) The particle accelerates because it loses potential energy as it moves from high to low potential.
B) The particle accelerates because it loses potential energy as it moves from low to high potential
C) The particle accelerates because it gains potential energy.
D) The particle accelerates because it moves towards negative charges.

Example 3-2

What is the electric potential at the edge of a hydrogen atom (a distance of 1 Å from the proton), if one sets 0 V at infinity? If an electron is located at a distance of 1 Å from the proton, how much energy is required to remove the electron; that is, how much energy is required to ionize the hydrogen atom?

Solution

We can easily calculate the electric potential, a distance of 1 Å from a proton, since this corresponds to the potential from a point charge (with $C = 0$):

$$V(\vec{r}) = \frac{kQ}{r} = \frac{(9\times 10^{9}\,\mathrm{N\cdot m^2/C^2})(1.6\times 10^{-19}\,\mathrm{C})}{(1\times 10^{-10}\,\mathrm{m})} = 14.4\,\mathrm{V}$$

We can calculate the potential energy of the electron (relative to infinity, where the potential is 0 V, since we chose $C = 0$):

$$U = (-e)V = (-1.6\times 10^{-19}\,\mathrm{C})(14.4\,\mathrm{V}) = -14.4\,\mathrm{eV} = -2.3\times 10^{-18}\,\mathrm{J}$$

where we also expressed the potential energy in electron volts. In order to remove the electron from the hydrogen atom, we must exert a force (do work) until the electron is infinitely far from the proton. At infinity, the potential energy of the electron will be zero (by our choice of $C = 0$). When moving the electron from the hydrogen atom to an infinite distance away, we must do positive work to counter the attractive force from the proton. The work that we must do is exactly equal to the change in potential energy of the electron (and equal to the negative of the work done by the force exerted by the proton):

$$W = \Delta U = (U_{final} - U_{initial}) = (0\,\mathrm{J} - -2.3\times 10^{-18}\,\mathrm{J}) = 2.3\times 10^{-18}\,\mathrm{J}$$

The positive work that we must do, exerting a force that is opposite to the electric force, is positive and equal to 2.3×10^{-18} J, or 14.4 eV. If you look up the ionization energy of hydrogen, you will find that it is 13.6 eV, so that this very simplistic model is quite accurate (we could improve the model by adjusting the proton-electron distance so that the potential is $13.6\,\text{V}$).

Discussion: In this example, we determined the electrical potential energy of an electron in a hydrogen atom, and found that it is negative, when potential energy is defined to be zero at infinity. In order to remove the electron from the atom, we must do positive work in order to increase the potential energy of the electron from a negative value to zero (the potential energy at infinity). This is analogous to the work that must be done on a satellite in a gravitationally bound orbit for it to reach escape velocity.

Example 3-3

Two large parallel plates are separated by a distance, L. The plates are oppositely charged and carry the same magnitude of charge per unit area, σ. What is the potential difference between the two plates? Write an expression for the electric potential in the region between the two plates. Assume that the plates are large enough that you can treat them as infinite (that is, neglect what happens near the edges).

Solution

Figure 3.4 shows a diagram of the two parallel plates with surface charge on them.

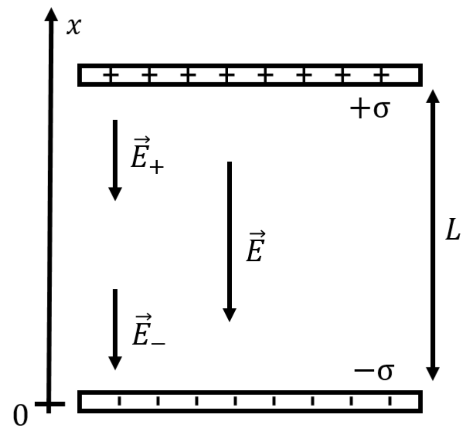

Figure 3.4: Two parallel plates with equal and opposite surface charge densities. In the region between the plates, the electric field is uniform.

We know from the previous chapters that the electric field from the positive plate does not depend on distance from the plate and is given by:

$$\vec{E}_+ = -\frac{\sigma}{2\epsilon_0}\hat{x}$$

if we approximate the plate as being infinitely large. This is a reasonable approximation for most points except those near the edges of the plate, which we ignore. The electric field from the negative plate will have the same magnitude and direction, so that the total electric field, $\vec{E}$, everywhere between the two parallel plates (as long as we are not near the edges) is given by:

$$\vec{E} = -\frac{\sigma}{\epsilon_0}\hat{x}$$

Note that the electric field outside the region between the two plates is zero everywhere, as the field from the positive and negative plates point in opposite directions outside the plates and thus cancel (except near the edges of the plates). For example, below the negative plate, the field from the negative plate points in the positive x direction (towards the negative plate), whereas the field from the positive plate points in the positive x direction (towards the positive plate).

We can now determine the potential difference between the two plates, since we know the electric field in that region. Using the coordinate system that is shown, we calculate the potential difference between the positive plate located at $x = L$ and the negative plate located at $x = 0$:

$$\Delta V = V(L) - V(0) = -\int_0^L \vec{E}\cdot d\vec{x} = -\int_0^L \frac{-\sigma}{\epsilon_0}\hat{x}\cdot d\vec{x} = \frac{\sigma}{\epsilon_0}\int_0^L dx = \frac{\sigma}{\epsilon_0}L$$

where we recognized that $\hat{x}$ and $d\vec{x}$ are parallel. It is very easy to get the wrong sign when calculating potential differences, so be careful!

Since the potential difference, $\Delta V = V(L) - V(0)$, is positive, the plate at $x = L$ is at a higher electric potential than the plate at $x = 0$. This makes sense, as a positive charge at rest would move from the positive plate to the negative plate, thus decreasing its potential energy, which corresponds to moving from a region of high electric potential to a region of low electric potential. Conversely, a negative charge at rest would move from the negative plate to the positive plate, decreasing its potential energy, but moving from a region of low electric potential to a region of high electric potential.

In general, if the electric field is constant, the change in potential between two points separated by a distance, L, along an axis that is anti-parallel with the field (in this example, the field points in the negative x direction) is given by:

$$\Delta V = -\int_0^L \vec{E}\cdot d\vec{x} = E\int_0^L dx = EL$$

Note that we can only calculate the difference in electric potential between plates, not the actual value of the potential, V. If we want to define a specific value of electric potential, we need to choose a location where we define $0\,\mathrm{V}$ to be. By convention, when possible, one chooses the negative plate to be the location of $0\,\mathrm{V}$. In order to determine the electric potential anywhere between the two plates, we can calculate the potential difference between the plate at $x = 0$ (the one at $0\,\mathrm{V}$) and some position between the plates along the x axis ($x < L$):

$$\Delta V = V(x) - V(0) = -\int_0^x E\hat{x} \cdot d\vec{x} = Ex = \frac{\sigma}{\epsilon_0} x$$
$$\therefore V(x) = V(0) + Ex = Ex = \frac{\sigma}{\epsilon_0} x$$

where we find that the electric potential increases **linearly** between its value at the negative plate (0 V) and its value at the positive plate (EL). Of course, we could have chosen any value of the electric potential for the negative plate, which is equivalent to choosing the value of the arbitrary constant, C.

In general, we can write the electric potential in a region of constant electric field, $\vec{E} = -E\hat{x}$, as:

$$V(x) = Ex + C$$

This scenario is very similar to the gravitational force near the surface of the Earth, where the gravitational field is constant. If you choose to define zero gravitational potential energy at the surface of the Earth, then, as you move up a distance h from the ground, your gravitational potential energy increases linearly with h ($U(h) = mgh$). In our case, we defined zero electrical potential energy to correspond to the location of the negative plate (the negative plate is thus like the surface of the Earth, with a constant electric field pointing towards it). As a positive charge moves a distance h away from the negative plate, it gains electric potential energy, $U(h) = qV(h) = qEh$, linearly with distance from the plate. If we release that positive charge, it will "fall" back onto the negative plate. The main difference with gravity, is that we can also have negative charges, which under gravity, would be similar to "negative masses" (it's not a thing), which would "fall upwards" (towards the positive plate).

Discussion: In this example, we examined the electric field between two parallel plates with opposite charges on them, and saw that the field is constant and uniform between the plates and zero outside (except for a small region near the edge of the plates where the assumption of infinitely large plates breaks down). We found that the electric potential decreases linearly as a function of distance from one of the plates. Because the electric field is constant between the two plates, the electric force on a charge can be treated in a similar way as the gravitational force on a mass near the surface of the Earth. The resulting electric potential is linear in the distance from the negative plate, just as mgh is linear in h, the distance to the surface of the Earth. Parallel plates are often used to accelerate charges, so they are useful to understand.

Checkpoint 3-5

If we defined a gravitional potential, V(h), for particles a small distance, h, from the surface of the Earth, it would have the form:

A) $V(h) = mgh + C$
B) $V(h) = gh + C$
C) $V(h) = mg + C$
D) $V(h) = -mgh + C$

3.2.2 Electric field from electric potential

Review Topics

- Section **??** on determining force from potential energy.
- Section **??** on gradients.

In the previous section, we found that we could determine the electric potential (a scalar) from the electric field vector. In this section, we show how to do the reverse, and determine the electric field vector from the electric potential. Consider, first, a one-dimensional case, where the electric field, $\vec{E}(x) = E(x)\hat{x}$, point in the x direction and depends on position, x. In this one-dimensional case, the electric potential is obtained from the negative anti-derivative of the electric field:

$$V(x) = -\int \vec{E}(x) \cdot d\vec{x} = -\int E(x)dx$$

The electric field must then be given by the negative of the derivative of the electric potential function:

$$\vec{E}(x) = -\frac{dV(x)}{dx}\hat{x}$$

Note that we can tell from the above that the electric field must have dimensions of electric potential over distance. The most common S.I. unit used to describe the electric field is V/m (Volts per meter).

This result is very similar to that obtained in Section **??**, where we examined how one could use the scalar potential energy, $U(x,y,z)$, to determine the vector for the force associated with that potential energy. The same holds for the electric force, where we can determine the electric force vector, $\vec{F}$, from the electric potential energy, and similarly the electric field from the electric potential. In three dimensions, if we know the electric potential energy as a function of position, $U(\vec{r}) = U(x,y,z)$, then the electric force vector is given by:

$$\vec{F}(x,y,z) = -\nabla U = -\frac{\partial U}{\partial x}\hat{x} - \frac{\partial U}{\partial y}\hat{y} - \frac{\partial U}{\partial z}\hat{z}$$

Similarly, but using force per unit charge (i.e. electric field) and potential energy per unit charge (i.e. electric potential), we find:

$$\vec{E}(x,y,z) = -\nabla V = -\frac{\partial V}{\partial x}\hat{x} - \frac{\partial V}{\partial y}\hat{y} - \frac{\partial V}{\partial z}\hat{z}$$

where, as you recall, ∇V, is called the gradient of the scalar field, $V(x,y,z)$. The gradient is a vector that points in the direction of maximal increase of the value of $V(x,y,z)$. For a positive charge, this corresponds to the direction of maximal increase in potential energy. A positive charge will experience a force in the opposite direction (in the direction where the potential energy decreases the fastest), and the electric field is thus in the opposite direction from the gradient of the electric potential.

3.2.3 Equipotential surfaces

We can visualize electric potential in several ways, since it is a scalar field (it has a single value that can differ everywhere in space). Figure 3.5 shows the electric potential near a positive charge, $+Q$, where one has chosen 0 V to be located at infinity. The left panel shows the electric potential as a "surface plot", where the vertical direction is the value of the electric potential. The right panel shows a "heat map" of the electric potential, where the colour corresponds to the value of the electric potential.

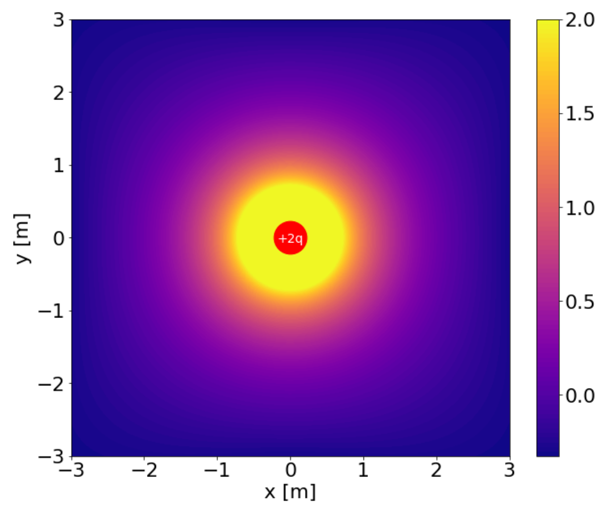

Figure 3.5: Electric potential heat map of a single positive charge.

The most common way to visualize the electric potential is to draw "contour lines", similar to how one draws contour lines on a geographical map. On a geographical map, contours correspond to lines of constant altitude, which are also lines of constant gravitational potential energy. Similarly, we can draw lines of constant electric potential to visualize the electric potential. Lines of constant potential are called "equipotential lines". In general, in three dimensions, regions of constant electric potential can be surfaces or volumes, called "equipotential surfaces/volumes". In Example 3-3 (with the parallel plates) each of the plates forms an equipotential surface (e.g. the electric potential was fixed to 0 V everywhere on the negative plate).

Recall that, at some point in space, the electric field vector always points in the opposite direction of the gradient of the electric potential. Namely, the electric field points in the direction in which the electric potential decreases the fastest. That direction must be perpendicular to the direction in which the electric potential does not change; in other words, the electric field vector is always perpendicular to equipotential lines/surfaces. More intuitively, one can think about a charge moving along an equipotential. By definition, the electric potential energy of the charge does not change if its moves along an equipotential. As a result, the electric force/field cannot do any work on the charge, and must thus be perpendicular to the path of the charge (which we chose to be an equipotential).

Conducting materials are always equipotential surfaces (or volumes) if charges are not moving inside the conductor. The electric field inside a conductor is always zero (in electrostatics, when charges are not moving), and thus, a charge moving through a conductor experiences no electric force and its electrical potential energy will be constant; in other words, the entire conductor is an equipotential. Similarly, because the electric field must always be perpendicular to an equipotential, electric field lines are always perpendicular to the surface of a conductor (in electrostatics).

In order to draw equipotential lines, one can start by drawing electric field lines, and then draw (closed) contour lines that are everywhere perpendicular to the electric field lines. This is illustrated in Figure 3.6.

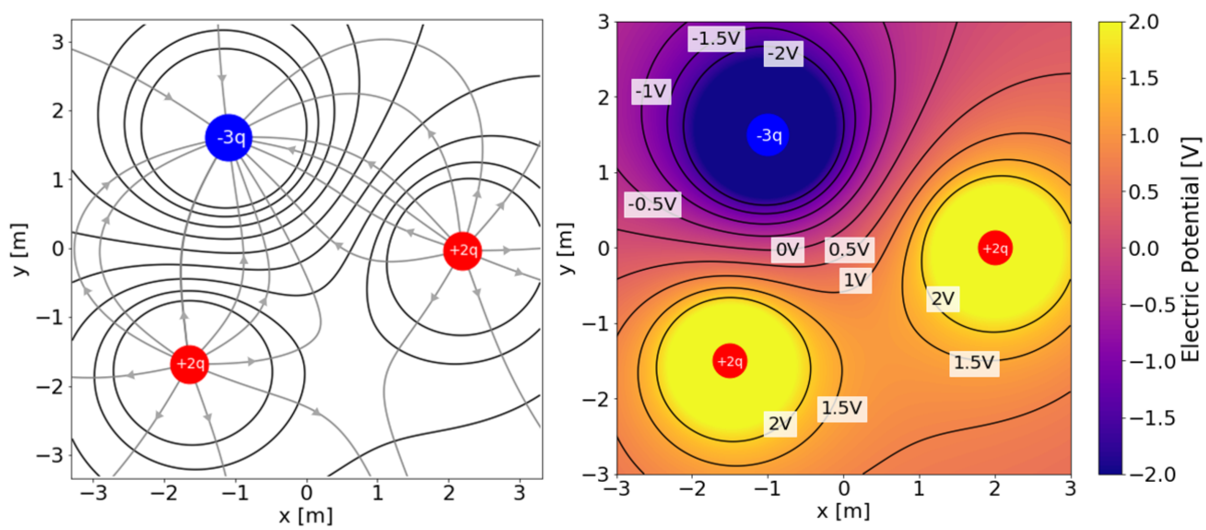

Figure 3.6: The electric field and equipotential lines caused by two $+2q$ charges and one $-3q$ charge (left) and its corresponding electric potential heatmap (right).

In general, it is preferable to draw equipotential lines that are separated by equal increments in electric potential (just as on a geographical map, the contour lines correspond to constant increments in altitude). This requires knowing a functional form for the electric potential. For example, the equipotential lines for a point charge located at the origin consist in concentric circles centred at the origin (in three dimensions, this results in concentric spherical equipotential surfaces). If we define 0 V to be at infinity, the electric potential is given by:

$$V(r) = \frac{kQ}{r}$$

In order to draw equipotential lines every, say, 10 V, the radii of the corresponding equipotential circles, for $V = 10\,\text{V}$, $V = 20\,\text{V}$, $V = 30\,\text{V}$, etc., are given by:

$$r = \frac{kQ}{V}$$
$$r_{10V} = \frac{kQ}{(10\,\text{V})} \quad r_{20V} = \frac{kQ}{(20\,\text{V})} \quad r_{30V} = \frac{kQ}{(30\,\text{V})} \quad \dots$$

3.3 Calculating electric potential from charge distributions

In this section, we give two examples of determining the electric potential for different charge distributions. We have two methods that we can use to calculate the electric potential from a distribution of charges:

1. Model the charge distribution as the sum of infinitesimal point charges, dq, and add together the electric potentials, dV, from all charges, dq. This requires that one choose $0\,\text{V}$ to be located at infinity, so that the dV are all relative to the same point.
2. Calculate the electric field (either as a integral or from Gauss' Law), and use:

$$\Delta V = V(\vec{r}_B) - V(\vec{r}_A) = -\int_A^B \vec{E} \cdot d\vec{r}$$

The first method is similar to how we calculated the electric field for distributed charges in chapter **??**, but with the simplification that we only need to sum scalars instead of vectors. The second method was already introduced in this chapter.

Example 3-4

A ring of radius R carries a total charge $+Q$. Determine the electric potential a distance a from the centre of the ring, along the axis of symmetry of the ring. Assume that zero electric potential is defined at infinity.

Solution

Figure 3.7 shows a diagram of the ring, and our choice of infinitesimal charge, dq.

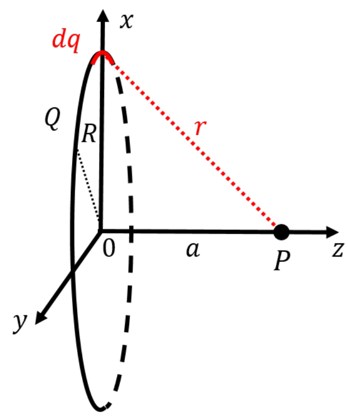

Figure 3.7: Determining the electric potential on the axis of a ring of radius R carrying charge Q.

In order to calculate the electric potential at point, P, with 0 V defined to be at infinity, we first calculate the infinitesimal potential at P from the infinitesimal point charge, dq:

$$dV = k\frac{dq}{r}$$

The total electric potential is then the sum (integral) of these potentials:

$$V = \int dV = \int k\frac{dq}{r} = \frac{k}{r}\int dq = k\frac{Q}{r} = k\frac{Q}{\sqrt{a^2 + R^2}}$$

where we recognized that k and r are the same for each dq, so that they could factor out of the integral. $\int dq = Q$ is then just the sum of the infinitesimal charges, which must add to the charge of the ring.

Discussion: In this example, we determined the electric potential, relative to infinity, a distance a from the centre of a charge ring, along its axis of symmetry. We modelled the ring as being made of many infinitesimal point charges, and summed together the infinitesimal electric potentials from those charges relative to infinity. This was much simpler than determining the electric field, since electric potential is a scalar and we do not need to consider how the components from different dq along the ring will cancel.

Example 3-5

A long, thin, straight wire carries uniform charge per unit length, λ. The electric potential difference between points located at distances $r_B = 2\,\text{cm}$ and $r_A = 1\,\text{cm}$ from the wire is found to be $V(r_B) - V(r_A) = -100\,\text{V}$. What is the linear charge density on wire, λ?

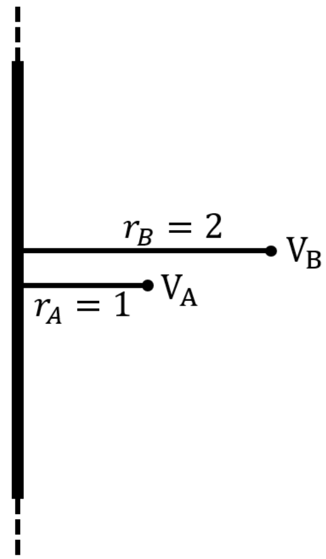

Figure 3.8: A long thin wire with measurements of electric potential at varying points

Solution

In this case, we can use Gauss' Law to determine the electric field at a certain distance from the wire. From that, we can calculate the electric potential difference between any two points near the wire, and thus the charge density on the wire.

By using a cylindrical surface of length, L, and radius, r, we can use Gauss' Law to determine the field at a distance, r, from the wire:

$$\oint \vec{E}\cdot d\vec{A} = \frac{Q^{enc}}{\epsilon_0}$$
$$2\pi rLE = \frac{\lambda L}{\epsilon_0}$$
$$\therefore \vec{E}(r) = \frac{\lambda}{2\pi\epsilon_0 r}\hat{r}$$

Using the electric field, we can calculate the potential difference between two points that are at distances, r_A and r_B, from the wire:

$$\begin{aligned}
\Delta V = V(r_B) - V(r_A) &= -\int_{r_A}^{r_B} \vec{E}\cdot d\vec{r} \\
&= -\int_{r_A}^{r_B} \left(\frac{\lambda}{2\pi\epsilon_0 r}\hat{r}\right)\cdot d\vec{r} = -\frac{\lambda}{2\pi\epsilon_0}\int_{r_A}^{r_B}\frac{1}{r}\hat{r}\cdot d\vec{r} = -\frac{\lambda}{2\pi\epsilon_0}\int_{r_A}^{r_B}\frac{1}{r}dr \\
&= -\frac{\lambda}{2\pi\epsilon_0}\left[|\ln(r)|\right]_{r_A}^{r_B} = -\frac{\lambda}{2\pi\epsilon_0}\ln\left(\frac{r_B}{r_A}\right) \\
\therefore \Delta V &= \frac{\lambda}{2\pi\epsilon_0}\ln\left(\frac{r_A}{r_B}\right)
\end{aligned}$$

where, in the second last line, we removed the absolute value from the logarithm, since $r_A < r_B$, and in the last line, we removed the minus sign by inverting the argument of the logarithm. Since we know the potential difference, ΔV, for two points located at distances $r_B = 2\,\text{cm}$ and $r_A = 1\,\text{cm}$, we can determine the charge density on the wire:

$$\begin{aligned}
\Delta V &= V(r_B) - V(r_A) = -100\,\text{V} \\
\Delta V &= \frac{\lambda}{2\pi\epsilon_0}\ln\left(\frac{r_A}{r_B}\right) \\
\therefore \lambda &= \frac{2\pi\epsilon_0\Delta V}{\ln\left(\frac{r_A}{r_B}\right)} = \frac{2\pi(8.85\times10^{-12}\,\text{C}^2\cdot\text{N}^{-1}\cdot\text{m}^{-2})(-100\,\text{V})}{\ln\left(\frac{1}{2}\right)} = 8.02\times10^{-9}\,\text{C/m}
\end{aligned}$$

where, again, one needs to be very careful with the signs! Note that it also makes sense that the potential difference, $\Delta V = V(r_B) - V(r_A)$, is negative, since r_A is closer to the positively charged wire. A positive charge at rest would move away from the positively charged wire, from r_A to r_B, from high potential to low potential.

Discussion: In this example, we showed how to determine the electric potential near an infinitely long charged wire by using the electric field that we determined from Gauss' Law. By knowing the potential difference between two points near the wire, we were then able to infer the charge density on the wire.

3.4 Electric field and potential at the surface of a conductor

If we consider a spherical conductor with charge, $+Q$, and radius, R, the electric field at the surface of the sphere is given by:

$$E = k\frac{Q}{R^2}$$

as we found in the Chapter 2. If we define electric potential to be zero at infinity, then the electric potential at the surface of the sphere is given by:

$$V = k\frac{Q}{R}$$

In particular, the electric field is related to the potential at the surface by:

$$E = \frac{V}{R}$$

Thus, if two spheres are at the same electric potential, the one with the smaller radius will have a stronger electric field at its surface.

Because a conducting sphere is symmetric, the charges will distribute themselves symmetrically around the whole surface of the sphere. The charge per unit area, σ, at the surface of the sphere is thus given by:

$$\sigma = \frac{Q}{4\pi R^2}$$

The charge density can be related to the electric field at the surface of the sphere:

$$E = k\frac{Q}{R^2} = k\frac{4\pi R^2\sigma}{R^2} = 4\pi\sigma k = \frac{\sigma}{\epsilon_0}$$

where in the last equality, we used replaced k with ϵ_0 and confirmed the general result from Section 2.3, where we determined the electric field near a conductor with surface charge, σ.

Consider a sphere of radius, R_1, that carries total charge, $+Q$. A neutral second, smaller, conducting sphere, of radius R_2 is then connected to the first sphere, using a conducting wire, as in Figure 3.9.

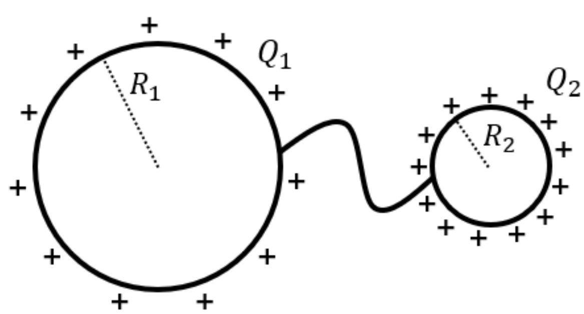

Figure 3.9: Two conducting spheres are connected by a conducting wire. The charge Q that was originally on the larger sphere distributes itself onto the two spheres.

Because the charges on the large sphere can move around freely, some of them will move to the smaller sphere. Very quickly, the charges will stop moving and the spheres of radius, R_1 and R_2, will end up carrying charges, Q_1 and Q_2, respectively (we assume that the wire is small enough that negligible amounts of charge are distributed on the wire). Since the two conducting spheres are connected by a conductor, they form an equipotential, and are thus at the same voltage, V, relative to infinity. Since the two spheres are at the same electric potential, the electric field at the surface of each sphere are related:

$$E_1 = \frac{V}{R_1}$$
$$E_2 = \frac{V}{R_2}$$
$$\therefore \frac{E_2}{E_1} = \frac{R_1}{R_2}$$
$$\therefore E_2 = E_1 \frac{R_1}{R_2}$$

and the electric field at the surface of the smaller sphere, E_2 is stronger since $R_2 < R_1$. We can also compare the surface charge densities on the two spheres:

$$E_1 = \frac{\sigma_1}{\epsilon_0}$$
$$E_2 = \frac{\sigma_2}{\epsilon_0}$$
$$\therefore \frac{\sigma_2}{\sigma_1} = \frac{E_2}{E_1} = \frac{R_1}{R_2}$$
$$\therefore \sigma_2 = \sigma_1 \frac{R_1}{R_2}$$

and we find that the charge density is higher on the smaller sphere. Thus, there are more charges per unit area on the smaller sphere than the bigger sphere.

We can generalize this model to describe charges on any charged conducting object. If charges are deposited on a conducting object that is not a sphere, as in Figure 3.10, they will not distribute themselves uniformly. Instead, there will be a higher charge density (charges per unit area), near parts of the object that have a small radius of curvature (sharp points on the object in particular), just as the charge density was higher on the smaller sphere described above. As a consequence of the higher concentration of charges near the "pointier" parts of the object, the electric field at the surface will be the strongest in those regions (as it is stronger at the surface of the smaller sphere described above).

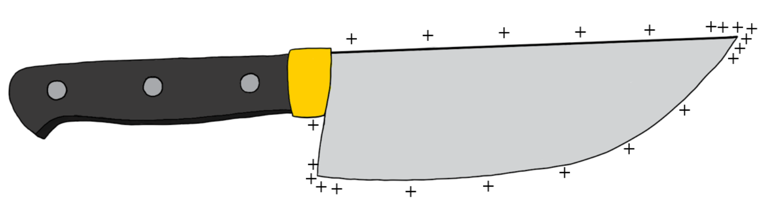

Figure 3.10: On an uneven conductor, charges will accumulate on the sharper points, where the radius of curvature is the smallest.

In air, if the electric field exceeds a magnitude of approximately $3 \times 10^6\,\mathrm{V/m}$, the air is said to "electrically breakdown". The strong electric field can remove electron from atoms in the air, ionizing the air and making it conductive. Thus, if the electric field at a point on the surface of a conductor is very strong, the air near that point will break down, and charges will leave the conductor, through the air, to find a location with lower electric potential energy (usually the ground). Electric breakdown is what we experience as a spark (or lightning, on a larger scale), and is usually a discrete (and potentially dramatic) event. Corona discharge is another mechanism whereby the strong electric field can make the air conductive, but in this case charges leak into the air more gradually, unlike in the case of electrical break down. Charges leaking into air through Corona discharge will emit a faint blueish light (the "Corona") as well as an audible hissing sound.

Objects that are designed to hold a high electric potential (for example the electrodes on high voltage lines) are usually made very carefully so that they have a very smooth surface and no sharp edges. This reduces the risk of breakdown or corona discharge at the surface which would result in a loss of charge.

Contrary to popular belief, lightning rods are not designed to attract lightening. Instead, lightning rods are designed to be conductors with a very sharp point, so that corona discharge can occur at their tip. This allows charges to slowly leak off from the Earth into the cloud through Corona discharge, thereby reducing the potential difference between the cloud and Earth so that a lightning strike (electrical breakdown) does not occur. When a lightning strike does occur, it will hit the lightning rod, since the electric field at the top of the rod is high and that is the most likely point for the air to break down; but, that is not the goal of the lightning rod!

3.5 Capacitors

Capacitors are common electronic devices that are used to store electric charge for a variety of applications. A capacitor is usually constructed with two conducting plates (called "terminals" or "electrodes") separated by either air or an insulating material.

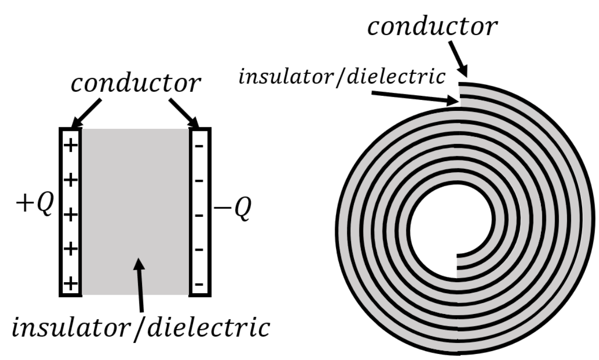

Figure 3.11: Two examples of capacitors. The left panel shows a "parallel plate" capacitor, and the right panel shows a cylindrically shaped capacitor obtained by "rolling up" a parallel plate capacitor.

Figure 3.11 shows two examples of capacitors. The left panel shows a "parallel plate" capacitor, consisting of two conducting plates separated by air or an insulator. The plates are conducting in order for one to be able to easily add and remove charge to the plates. The plates always hold equal and opposite charges. The right panel shows a more practical implementation of a capacitor that could be used in a circuit, which is simply made by "rolling up" a parallel plate capacitor (with an insulator instead of air separating the plates so that they do not touch).

3.5.1 Capacitance

As long as the quantities of charge involved are not too large, it has been observed that the amount of charge, Q, that can be stored on a capacitor[1], is linearly proportional to the potential difference, ΔV, between the two plates:

$$Q \propto \Delta V$$

$$\boxed{Q = C\Delta V}$$

The constant of proportionality, C, between charge and potential difference across the capacitor (usually called voltage across the capacitor) is called "capacitance", and has S.I. units of "Farads", F. The capacitance of a particular capacitor is a measure of how much charge it can hold at given voltage and depends on the geometry of the capacitor as well as the material between the terminals. If too much charge is placed on a capacitor, the material between the two plates will break down, and a spark will usually damage the capacitor as well as discharge it.

We can easily calculate the capacitance of a parallel plate capacitor. We model the capacitor as being made of two conducting plates, each with area, A, separated by a distance, L, and holding charge with magnitude, Q. The surface charge density on one of the plates, σ, is just given by:

$$\sigma = \frac{Q}{A}$$

In Example 3-3, we found an expression for the potential difference between two parallel plates:

$$\Delta V = \frac{\sigma}{\epsilon_0}L = \left(\frac{L}{A\epsilon_0}\right)Q$$

Comparing with, $Q = C\Delta V$, the capacitance of the parallel plate capacitor is found to be:

$$C = \epsilon_0\frac{A}{L}$$

It makes sense that the capacitance, the amount of charge that can be stored at a given voltage, increases if the plates have a larger area (more space for charges), and decreases if the plates are further apart (smaller electric field).

Capacitors are used in many touch screens. For example, these might be made of glass (an insulator), with a thin metal coating that one touches to interact with the screen (one of the plates). As you touch the metal plate, you effectively change the capacitance of the screen, which can be sensed and modelled to determine the location of your finger(s). Modern touch screen have many capacitors built directly into the screen, and function based on this principle.

[1]This is the amount of charge on one of the plates. As a whole, the capacitor is neutral.

Checkpoint 3-6

A capacitor holds 0.2C of charge when it has a potential difference of 500V between its plates. If the same capacitor holds 0.15C of charge, what is the potential difference between its plates?

A) 375 V
B) 500 V
C) 75 V
D) 150 V

3.5.2 Dielectric materials

In practice, capacitors always have an insulating material between the two plates. The material is chosen to have a higher breakdown voltage than air, so that more charges can be stored before a breakdown occurs. It has also been experimentally observed that the capacitance increases with certain materials, so called "dielectric materials". A dielectric material has a "dielectric constant", K, defined to be the amount by which the capacitance increases:

$$C = KC_0$$

where C is the capacitance with the material in place, and C_0 is the capacitance when there is vacuum between the plates (the dielectric constant of air is very close to 1). Often, rather than the dielectric constant, one uses the "permittivity", ϵ, of a material:

$$\epsilon = K\epsilon_0$$

based on the permittivity of free space, ϵ_0. The capacitance of a parallel plate capacitor, with a material that has permittivity, ϵ, is thus given by:

$$C = K\epsilon_0\frac{A}{L} = \epsilon\frac{A}{L}$$

Dielectrics materials are made of molecules that can be polarized (as water), namely molecules that have a non-zero electric dipole moment. When the dielectric material is placed between the plates, the dipoles inside the material align themselves with the electric field from the plates. This leads to a second electric field, from the dipoles, in the opposite direction of the field from the plates, thus reducing the total electric field between the plates. This, in turn, allows more charges to be held on the plate for a given voltage. This is illustrated in Figure 3.12

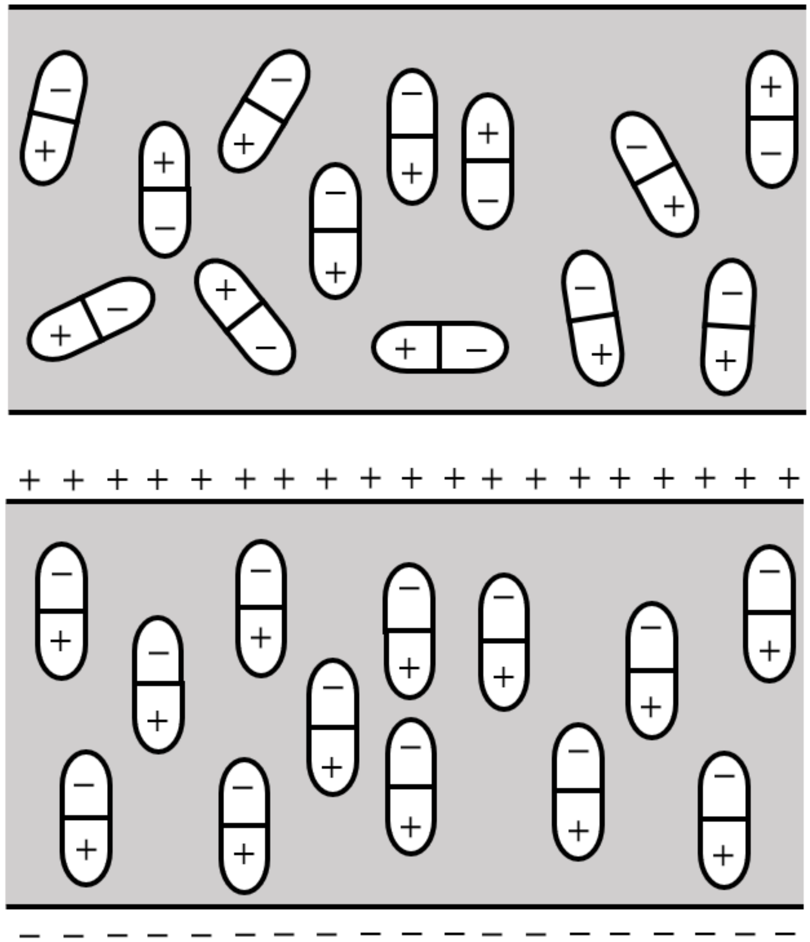

Figure 3.12: Unpolarized dipoles between two charged plates (top) and polarized dipoles between two charged plates (bottom).

Note that, in a dielectric material with permittivity, ϵ, Gauss' Law is modified to read:

$$\oint \vec{E} \cdot d\vec{A} = \frac{Q^{enc}}{\epsilon}$$

where the permittivity of free space, ϵ_0, is simply replaced with the permittivity of the material, ϵ.

3.5.3 Energy stored in a capacitor

The charges stored on a capacitor have electrical potential energy: if one were to place a conductor between the plates, charges would immediately conduct from one plate to the other and gain kinetic energy. We can model the amount of energy stored on the capacitor by considering how much work it takes to place the charges on the capacitor.

Imagine that both plates on the capacitor start with a charge of magnitude, q. We then remove an infinitesimal negative charge, with magnitude dq, from the positive plate and place it on the negative plate. This required work, since we had to pull this negative charge away from the positive plate. If the potential difference across the plates is ΔV, then we had to do an amount of work given by:

$$dW = Vdq$$

since the charge dq has now gained potential energy $\Delta V dq$. The potential difference is however dependent on the (constant) capacitance of the capacitor, and the amount of charge, q, already stored on the plates:

$$\begin{aligned} q &= C\Delta V \\ \therefore \Delta V &= \frac{q}{C} \end{aligned}$$

In order to determine the work required to transfer a total amount of charge, Q, we sum the work in transferring each infinitesimal charge, dq:

$$W = \int dW = \int_0^Q Vdq = \int_0^Q \frac{q}{C}dq = \frac{1}{2}\frac{Q^2}{C}$$

Thus, the total potential energy that is stored on a capacitor is given by:

$$U = \frac{1}{2}\frac{Q^2}{C} = \frac{1}{2}Q(\Delta V)^2 = \frac{1}{2}Q\Delta V$$

where we made use of $Q = C\Delta V$ to show the formula with different choices of variables. In either case, the amount of energy that is stored increases with the amount of charge, the capacitance, and the voltage across the capacitor. Capacitors are useful because this energy can be released quickly, as in the bright flash of light required for flash photography.

3.6 Summary

Key Takeaways

The electric force is conservative, so we can define a potential energy function, $U(\vec{r})$. The potential energy function for a point charge, q, at position, $\vec{r}$, relative to a point charge, Q, is given by:

$$U(\vec{r}) = \frac{kQ}{r}q + C$$

where, C, is an arbitrary constant, since only difference in potential energy are physically meaningful (as they correspond to work). Note that the value of the electrical potential energy will depend on the relative sign of q and Q.

If a collection of charges are held together, the total electrical potential energy that is stored is called "electrostatic potential energy".

In a similar way as the electric field, $\vec{E}(\vec{r})$, corresponds to electric force per unit charge, "electric potential", $V(\vec{r})$, corresponds to electric potential energy per unit charge. The electric potential at a position, $\vec{r}$, relative to a point charge, Q, is given by:

$$V(\vec{r}) = \frac{kQ}{r} + C$$

and also depends on an arbitrary constant, since only differences in electric potential will lead to differences in potential energy. The value of the electric potential, V, at some position in space, $\vec{r}$, allows us to determine the electric potential energy, U, at that position for any charge, q:

$$U = qV$$

This is analogous to determining the force on a charge q when we know the electric field at some point in space:

$$\vec{F} = q\vec{E}$$

Differences in electric potential are called "voltages", and the S.I. unit of potential is called the "volt" (V). In S.I. units, the electric field is often expressed in units of volts per metre (V/m).

When a particle with charge, q, changes position such that the corresponding change in electric potential is ΔV, the particle's potential energy will change by:

$$\Delta U = q\Delta V$$

In particular, a negative charge will experience a decrease in potential energy when the electric potential increases, whereas a positive charge will experience an increase in potential energy when the electric potential increases. This reflects the fact that the electric force associated with the electric potential will act in opposite directions on a positive and a negative charge.

In order to describe the energies of particles interacting with electric forces, it is more convenient to use the "electron volt" instead of the Joule. An electron volt is defined as the energy that is gained by a charge with a magnitude e (the magnitude of the charge of the electron) when accelerated through a potential difference of $\Delta V = 1V$:

$$1\,\text{eV} = (e)(1\,\text{V}) = 1.6 \times 10^{-19}\,\text{J}$$

The electric potential function can be determined in two different ways:

1. By modelling the charge distribution as the sum of infinitesimal point charges, dq, and adding together the electric potentials, dV, from all charges, dq. This requires that one choose $0\,\text{V}$ to be located at infinity, so that the dV are all relative to the same point.
2. By calculating the electric field (either as a integral or from Gauss' Law), and using:

$$\Delta V = V(\vec{r}_B) - V(\vec{r}_A) = -\int_A^B \vec{E} \cdot d\vec{r}$$

It is worth noting that one needs to be very careful with the signs when using the above integral. In particular note that one takes the negative of the integral, from A to B, to determine the potential at B minus the potential at A.

Similarly, one can determine the value of the electric field, $\vec{E}(\vec{r}) = \vec{E}(x, y, z)$, from the electric potential, $V(\vec{r}) = V(x, y, z)$:

$$\vec{E}(x, y, z) = -\nabla V = -\frac{\partial V}{\partial x}\hat{x} - \frac{\partial V}{\partial y}\hat{y} - \frac{\partial V}{\partial z}\hat{z}$$

where, ∇V, is the gradient of the electric potential.

The electric potential can be visualized in a number of ways. The most common is to draw contours of constant electric potential, akin to the contours on geographical maps that are used to show regions of constant altitude (i.e. constant gravitational potential energy).

Regions of constant electric potential are called "equipotentials", and can be lines, surfaces or volumes. Equipotentials are always perpendicular to the electric field. In electrostatics (when charges are not moving), the electric field in a conductor must be zero, so that a conductor always forms an equipotential, and the electric field at the surface of a conductor is always perpendicular to the surface.

When charges are placed on a conductor, they will spread out along the outer surface of the conductor. The surface density of charges will be the highest where the conductor has the smallest radius of curvature (e.g. at a sharp point). Consequently, the electric field at the surface of a charged conductor is highest near sharp points.

Capacitors are devices that are used to store charge. They are usually made using two conducting plates ("terminals" or "electrodes") that hold equal and opposite charge, Q, at a fixed potential difference, ΔV , between the electrodes. The amount of charge that is stored on the capacitor is observed to be proportional to the potential difference between the electrodes:

$$Q = C\Delta V$$

where the constant of proportionality, C, is called the "capacitance" of the capacitor. The S.I. unit of capacitance is the "Farad" (F). The capacitance of a capacitor depends on its geometry (e.g. its size) and the materials that it is placed between the electrodes.

Usually, one places a dielectric material between the two electrodes in order to increase the capacitance, and to reduce the risk of breakdown. If that material has a "dielectric constant", K, then the capacitance is given by:

$$C = KC_0$$

where, C_0, corresponds to the capacitance if there were vacuum between the electrodes. The dielectric constant of air is very close to 1, so that a capacitor in air is very similar to a capacitor in vacuum. A dielectric material is one that is made of molecules that can be polarized under the presence of an electric field; that is, the molecules have an electric dipole moment. When the molecules in a material are polarized, this reduces the total electric field in the material, which increases the capacitance of the capacitor. Inside a dielectric material, we can define the "permittivity", ϵ, as:

$$\epsilon = K\epsilon_0$$

where ϵ_0 is the permittivity of free space. Within a dielectric material, Gauss' Law is modified to:

$$\oint \vec{E} \cdot d\vec{A} = \frac{Q^{enc}}{\epsilon}$$

Since charges are held at a fixed potential difference on a capacitor, capacitors are a way of storing electric potential energy. The amount of electric potential energy stored in a capacitor with capacitance, C, when the capacitor has a potential difference, ΔV, across its electrodes, is given by:

$$U = \frac{1}{2}\frac{Q^2}{C} = \frac{1}{2}C(\Delta V)^2 = \frac{1}{2}Q\Delta V$$

Important Equations

Charge stored in a capacitor:

$$Q = C\Delta V$$

Electric potential energy

$$U = qV$$

Electric potential:

$$\Delta V = V(\vec{r}_B) - V(\vec{r}_A)$$

$$\Delta V = -\int_A^B \vec{E} \cdot d\vec{r}$$

Important Definitions

Electric Potential: A measurement of electric potential energy per unit charge. SI units: [V]. Common variable(s): V, often appearing as ΔV (potential difference).

Capacitance: A measurement which states how much charge a capacitor can hold given the potential difference between the plates of the capacitor. SI units: [F]. Common variable(s): C.

Dielectric constant: A constant which is defined as the ratio of the dielectric permittivity of a substance and the dielectric permittivity of a vacuum. SI units: none. Common variable(s): K.

3.7 Thinking about the material

Reflect and research

1. Explain how the capacitance can increase when a dielectric material is used.
2. Explain how a corona ring works.
3. Which types of electrodes are most common? Why?

To try at home

1. Try to release a static discharge from your finger to some metal object. Measure the distance between your finger and the metal object at the time of the discharge. Knowing the breakdown voltage of air, what was the potential difference between your finger and the metal object just before the discharge?

To try in the lab

1. Propose an experiment to measure the point at which various substances experience electric breakdown.
2. Propose an experiment to experimentally measure the vacuum permittivity constant (ϵ_0) using equation (are we able to reference equations in this textbook? would be useful here)

3.8 Sample problems and solutions

3.8.1 Problems

Problem 3-1: Consider a long cylinder of radius R and charge density ρ. If the electric potential at the surface of the cylinder is $V_S = 100\,\mathrm{V}$, then what is the electric potential inside and outside of the cylinder as a function of r? (Solution)

Problem 3-2: Consider two nested conducting cylinders of length L which are seperated by the distance between the points a and b, which are both measured radially from the centre of the nested cylinders, as shown in Figure 3.13. If the linear charge density of both cylinders is λ, what is the capacitance between points a and b? (Solution)

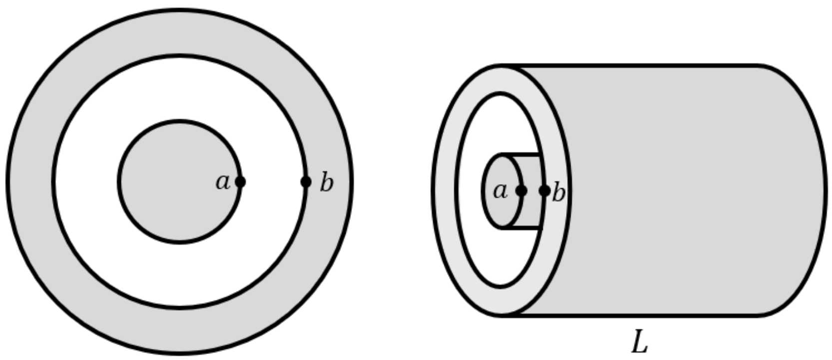

Figure 3.13: Two nested cylinders serpated by the distance between the points a and b.

3.8.2 Solutions

Solution to problem 3-1:

To calculate the electric potential inside and outside of the cylinder, we must first calculate the electric field. We will do this using Gauss' Law. We will define our enclosed charge using a cylinder of radius r which is less than R and length L:

$$\begin{aligned}
\int E dA &= \frac{Q_{enc}}{\epsilon_0} \\
E 2\pi r L &= \frac{\rho \pi r^2 L}{\epsilon_0} \\
E(r) &= \frac{\rho \pi r^2 L}{2\pi \epsilon_0 r L} \\
E(r) &= \frac{\rho r}{2\epsilon_0}
\end{aligned}$$

Calculating the electric field outside of the cylinder follows the same process:

$$\begin{aligned}
\int E dA &= \frac{Q_{enc}}{\epsilon_0} \\
E 2\pi r L &= \frac{\rho \pi R^2 L}{\epsilon_0} \\
E(r) &= \frac{\rho \pi R^2 L}{2\pi \epsilon_0 r L} \\
E(r) &= \frac{\rho R^2}{2\epsilon_0 r}
\end{aligned}$$

We now have the two equations which are required to calculate electric potential. To calculate electric potential, we will integrate from the starting point to r for each electric field function. We will begin with the potential difference inside of the cylinder:

$$\begin{aligned}
\Delta V &= -\int_r^R \vec{E} \cdot d\vec{r} \\
\Delta V &= -\int_r^R \frac{\rho r}{2\epsilon_0} \cdot d\vec{r} \\
\Delta V &= -\frac{\rho(R^2 - r^2)}{4\epsilon_0} \\
\Delta V &= \frac{\rho(r^2 - R^2)}{4\epsilon_0}
\end{aligned}$$

From here, we must solve for $V(r)$:

$$V(r) - V_S = \Delta V$$
$$V(r) - 100 = \frac{\rho(r^2 - R^2)}{4\epsilon_0}$$
$$V(r) = \frac{\rho(r^2 - R^2)}{4\epsilon_0} + 100$$

Which is our $V(r)$ inside of the shell. Calculating $V(r)$ outside of the shell follows the same process:

$$\Delta V = -\int_R^r \vec{E} \cdot d\vec{r}$$
$$\Delta V = -\int_R^r \frac{\rho R^2}{2\epsilon_0 r} \cdot d\vec{r}$$
$$\Delta V = -\frac{\rho R^2}{2\epsilon_0}(ln(r) - ln(R))$$
$$\Delta V = \frac{\rho R^2}{2\epsilon_0} ln(\frac{R}{r})$$

And now we solve for $V(r)$:

$$V(r) - V_S = \Delta V$$
$$V(r) - 100 = \frac{\rho R^2}{2\epsilon_0} ln(\frac{R}{r})$$
$$V(r) = \frac{\rho R^2}{2\epsilon_0} ln(\frac{R}{r}) + 100$$

Solution to problem 3-2:

To begin, we must apply Gauss' law to find the electric field for r where $a \geq r \geq b$. We know that the cylinder of radius a will carry the only enclosed charge between points a and b, so we can determine the following:

$$\Phi = \int \vec{E} d\vec{A} = \frac{Q_{enc}}{\epsilon_0}$$
$$E(2\pi r L) = \frac{\lambda L}{\epsilon_0}$$
$$E = \frac{\lambda}{2\pi\epsilon_0 r}$$

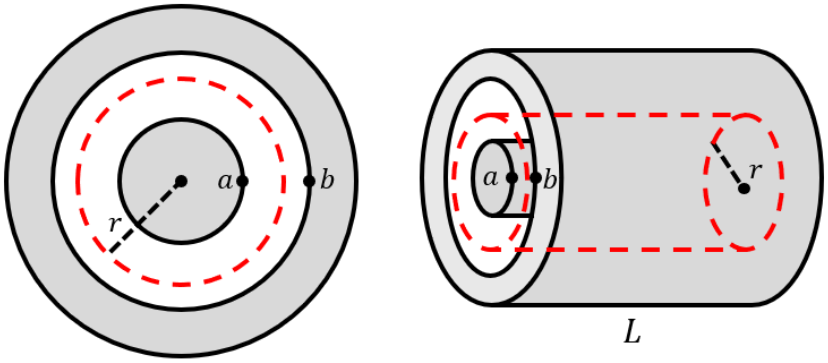

Figure 3.14: Solving for E between two cylinders using Gauss' law.

Now we must find ΔV. To do this, we will integrate E between the points a and b:

$$\Delta V = \int_a^b E dr$$
$$\Delta V = \int_a^b \frac{\lambda}{2\pi\epsilon_0 r} dr$$
$$\Delta V = \frac{\lambda}{2\pi\epsilon_0}(ln(b) - ln(a))$$
$$\Delta V = \frac{\lambda ln(\frac{b}{a})}{2\pi\epsilon_0}$$

From here, we know that $Q = C\Delta V$, so we can say that:

$$C\Delta V = \frac{\lambda}{L}$$
$$C = \frac{\lambda}{L\Delta V}$$
$$C = \frac{\lambda 2\pi\epsilon_0}{L\lambda ln(\frac{b}{a})}$$
$$C = \frac{2\pi\epsilon_0}{L ln(\frac{b}{a})}$$

4 Electric current

In this chapter, we introduce tools to model electric current, namely, the motion of charges inside a conductor. We will show how we can connect the microscopic motion of electrons to macroscopic quantities, such as current and voltage, that can be measured in the laboratory. We will also introduce the notion of resistance, as well as the resistor, a common component in electric circuits.

Learning Objectives

- Understand the differences in modelling conductors when charges are stationary or moving.
- Understand how to define current and current density.
- Understand the differences between resistance, resistivity, and conductivity.
- Understand Ohm's Law.
- Understand how to model how power is dissipated in a resistor.
- Understand how to model alternating current.
- Understand some elements of electrical safety.

Think About It

Why is it safe to touch the $300\,000\,\text{V}$ terminal of a Van de Graaf generator, and not the $12\,\text{V}$ terminal of a car battery?

A) The Van de Graaf generator cannot sustain a large current.
B) The Van de Graaf generator produces alternating current.
C) The car battery produces $12\,\text{V}$ of alternating voltage.

4.1 Current

In the preceding chapters, we examined "electrostatic" systems; those for which charges are not in motion. In electrostatic systems, the electric field inside of a conductor is zero (by definition, or charges would be moving, since they are free to move in a conductor). We argued that if charges are deposited onto a conductor, they would quickly arrange themselves into a static configuration (on the surface of the conductor).

Instead, we can build systems were charge move in a conductor. If we apply a fixed potential difference across a conductor, this will result in an electric field inside the conductor and the charges within will move as a result. In general, this requires that there be some sort of circuit formed, whereby charges enter one end of the conductor and exit the other. The most simple circuit that one can construct is to connect the two terminals of a battery to the ends of a conductor, as illustrated in Figure 4.1.

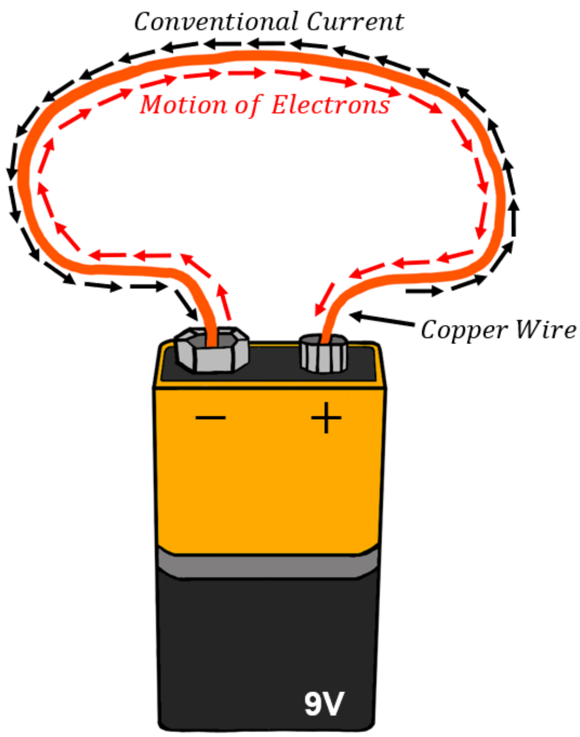

Figure 4.1: A simple circuit is created by connecting the terminals of a battery to a conducting material. Note that while electrons flow from the negative to the positive terminal of the battery, conventional current is defined as if it were positive charges moving in the opposite direction.

A battery (as we will see in more detail in Section 5) is a device that provides a source of charges and a fixed potential difference. For example, a 9 V battery has two terminals with a constant voltage of 9 V between them.

"Electric current" is defined to be the rate at which charges cross a given plane (usually a plane perpendicular to some conductor through which we want to define the current). We define current, I, as the total amount of charge, ΔQ, that flows through any cross-section of the conductor during an amount of time, Δt:

$$\boxed{I = \frac{\Delta Q}{\Delta t} = \frac{dQ}{dt}}$$

where we take a derivative if the rate at which charges flow is not constant in time. The S.I. unit of current is the Ampère (A), which is a base unit (note that the Coulomb (charge), is a derived unit, $1\,\text{C} = 1\,\text{A}\cdot\text{s}$). Current is defined to be positive in the direction in which positive charges flow. In almost all cases, it is negative electrons that flow through a material; the current is defined to be in the opposite direction from which the actual electrons are flowing, as illustrated in Figure 4.1. To distinguish that the current is in the direction opposite to that of the flowing electrons, one sometimes uses the term "conventional current" to indicate that the current is referring to a flow of positive charges.

Note that the definition of electric current is very similar to the "flow rate", Q, that we defined as the volume flow of a liquid across a given cross-section (Section **??**). As we continue to develop our description of current, you will notice that there are many similarities between describing the flow of an incompressible fluid and describing the flow of charges in a conductor.

We think of current as a macroscopic quantity, something that we can easily measure in the lab. Current is a measure of the average rate at which charges are moving through the conductor, and not a measure of what is going on at a microscopic level. In order to model the motion of charges at the microscopic level, we introduce the "current density", $\vec{j}$:

$$\boxed{\vec{j} = \frac{I}{A}\hat{E}}$$

where, I, is the current that flows through a surface with cross-sectional area, A, and $\hat{E}$ is a unit vector in the direction of the electric field at the point where we are determining the current density. The current density allows us to develop a microscopic description of the current, since it is the electric current per unit area and points in the direction of the electric field at some position. Given the current density, $\vec{j}$, one can always determine the current through a surface with area, A, and normal vector, $\hat{n}$:

$$I = A(\vec{j} \cdot \hat{n})$$

If the current density changes over the surface, one must take an integral instead:

$$I = \int \vec{j} \cdot d\vec{A}$$

where $d\vec{A}$, is a surface element with area, A, and direction given by the normal to the surface at that point. The overall sign of the current will be determined by the direction of the flow of positive charges.

Example 4-1

Electric current flows through a conductor with a narrowing cross section, as illustrated in Figure 4.2. If the cross-sectional area the conductor is A_1 at one end, and A_2, at the other end, what is the ratio of the current densities, j_1/j_2, at the two ends of the conductor?

Solution

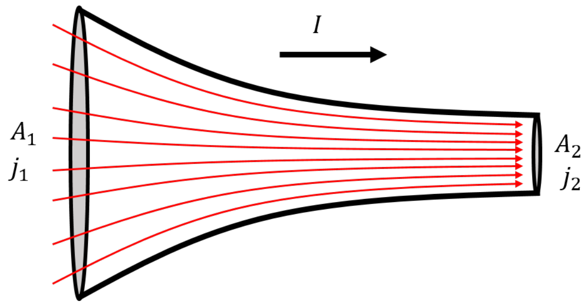

Figure 4.2: Current flows through a conductor with a cross-section that decreases from A_1 to A_2.

This situation is very similar to the flow of an incompressible fluid. In this case, the number of charges entering the conductor must be equal to the number of charges entering the conductor during a given amount of time. That is, the total current, I, must be the same at both ends, since there is no place in the conductor for charges to accumulate. Since the current must be the same on both ends, we can related the current densities at each end:

$$j = \frac{I}{A}$$
$$\therefore I = j_1 A_1 = j_2 A_2$$
$$\therefore \frac{j_2}{j_1} = \frac{A_1}{A_2}$$

and we find that the current density at the exit of the conductor must be higher than at the entrance. This is similar to the continuity equation in the Fluid Mechanics chapter (Section **??**), where the current density plays a role analogous to the velocity in the fluids case.

4.2 Microscopic model of current

Consider a cylindrical conductor of cross-sectional area, A, and length, L, as shown in Figure 4.3. A potential difference, ΔV, is applied across the length of the conductor, so that there is an electric field, $\vec{E}$, everywhere within the conductor. If the conductor were made of empty space, electrons would enter one end of the conductor, accelerate through the potential difference, and arrive at the other end with a high speed, having gained $e\Delta V$ of kinetic energy. In reality, the conductor is made of matter, and electrons do not accelerate continuously through the whole length of the conductor. Instead, they can only accelerate over a short distance before colliding with an atom in the material (rather, a tightly bound electron in the material), and losing their kinetic energy to the material, before accelerating again. The motion of electrons flowing in a conductor is illustrated in Figure 4.3 and shows electrons moving with a wide range of velocities following the collisions, and only an average motion in the direction anti-parallel to the electric field.

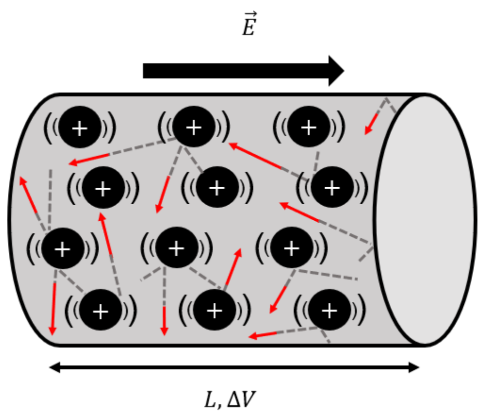

Figure 4.3: Electrons moving inside a conductor only "drift" on average in the direction anti-parallel to the electric field. In reality, they constantly collide with atoms in the material, transferring their kinetic energy into thermal energy of the conductor.

Thus, when the electrons arrive at the positive side of the conductor, they have not gained any kinetic energy. Instead, they have lost that kinetic energy to atoms of the conducting material through collisions; those atoms then vibrate which we can measure as an increase in temperature of the material. When current flows through a conductor, that conductor will heat up; this is how the heating elements in your toaster work!

We model the motion of electrons as charges "drifting" through the conductor with a velocity, $\vec{v}_d$, the "drift velocity", as illustrated in Figure 4.4. In reality, of course, the electrons are only moving on average with the drift velocity, and their instantaneous speed is generally much larger than the drift velocity and can be in any direction, as illustrated above.

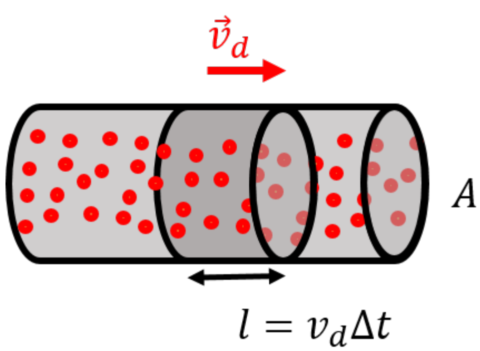

Figure 4.4: A section of electrons of length l drifting through a conductor of cross-sectional area, A.

In a conducting material, each atom will generally have one "free" electron that is loosely bound and able to easily move through the material. The number of free electrons available for conduction per unit volume, n, will depend on the density of the material. Consider, then, the motion of the conduction electrons present in a section of length, l, of a conductor, as illustrated in Figure 4.4. The amount of charge, ΔQ, contained in a section of the conductor with length, l, is given by:

$$\Delta Q = -enAl$$

where Al is the volume of that section of the conductor, and, e, is the magnitude of the charge of the electron. The negative sign is to indicate that the charges are negative (they are electrons). That charge will take an amount of time, Δt, to flow through a given plane of the conductor, so that we can relate the length of the section, l, to the drift speed and Δt:

$$l = v_d \Delta t$$

Thus, the current that flows through a cross-section of the conductor is given by:

$$I = \frac{\Delta Q}{\Delta t} = \frac{-enAl}{\Delta t} = -enAv_d$$
$$\therefore \boxed{I = -enAv_d}$$

which allows us to connect a macroscopic quantity, current, to the microscopic description of charges moving. Note that the negative sign reflects the fact that the current (of positive charges) is in the opposite direction from the drift velocity of the (negative) electrons. The current density is directly related to the microscopic quantities, since it does not depend on the (macroscopic) cross-sectional area, A, of the conductor:

$$\vec{j} = \frac{I}{A}\hat{E} = -en\vec{v}_d$$
$$\therefore \boxed{\vec{j} = -en\vec{v}_d}$$

where, again, the negative sign indicates that the current density is in the opposite direction from the actual drift velocity of the electrons, which itself is anti-parallel to the electric field.

Example 4-2

A current of 1 A is measured in a copper wire with a diameter of 1 mm. What is the drift velocity of the electrons? Assume that each atom of copper provides one "free electron" for conduction.

Solution

In order to determine the drift velocity of electrons, we need to know the density of free electrons in copper. To do this, we need to determine how many copper atoms there are per unit volume. The density of copper is $\rho = 8.92 \times 10^3\,\mathrm{kg/m^3}$ and the atomic mass unit of copper is 63.5 amu (1 mole of copper weighs 63.5 g). The number of copper atoms per unit volume is thus:

$$n = \frac{(6.022 \times 10^{23}\,\mathrm{mole^{-1}})(8.92 \times 10^3\,\mathrm{kg/m^3})}{(63.5 \times 10^{-3}\,\mathrm{kg/mole})} = 8.46 \times 10^{28}\,\mathrm{m^{-3}}$$

Since each copper atom contributes one free electron, this is the same as the density of free electrons. From this, we easily obtain the drift velocity, from the current:

$$\begin{aligned} v_d = \frac{j}{en} = \frac{I}{Aen} &= \frac{(1\,\mathrm{A})}{\pi(0.0005\,\mathrm{m})^2(1.6\times10^{-19}\,\mathrm{C})(8.46\times10^{28}\,\mathrm{m}^{-3})} \\ &= 9.4\times10^{-5}\,\mathrm{m/s} \sim 0.1\,\mathrm{mm/s} \end{aligned}$$

The drift velocity is thus very slow, less than one millimetre per second. Note that a copper wire would not actually be able to sustain such a high current density without damage.

Josh's Thoughts

There are a few types of velocities which can be easily confused when discussing current: fermi velocity, drift velocity, and the velocity at which a circuit is completed.

First, there is fermi velocity, which is defined as the velocity generated when an electron moves from the highest energy level in an atom to the lowest energy level in that atom. This process can be thought of as if the electron is moving through a discrete potential difference. The kinetic energy generated in the jump from high to low potential causes electrons in good conductors to travel at roughly 1/200 the speed of light.

While electrons do move at their fermi velocity in a conductor, they do not move in a uniform path through the conductor towards the end of the circuit. Most of an electron's movement in a wire is chaotic, but in a DC circuit, the electrons have a drift velocity. This drift velocity is defined as the net velocity of electrons in a conductor, and is caused by the applied electric field which has a small amount of influence on the direction of the quickly moving electron's motion. The drift velocity of electrons is very slow, often having a magnitude as small as 10s of microns per second.

When comparing drift velocity to fermi speed, imagine yourself standing in a large cylinder, which will represent the conductor in this analogy. The interior of this cylinder is lined with cannons that shoot rubber balls in all directions, which will be the electrons moving at their fermi velocity. Now, imagine that you are attempting to move these high-speed rubber balls from one end of the cylinder to the other by blowing a hair dryer in that direction, which is the electric field inducing a drift velocity.

Now that we understand the quantum chaos that occurs in a conductor, you may be thinking to yourself, "why does the light bulb turn on so quickly after I flick the light switch?". This is a reasonable thought, because we have only covered the motion of single particles in a conductor. When an electron moves very slightly (at its drift velocity), it will push other electrons in the conductor forward, causing a chain reaction of electrons pushing one another forward. This movement causes electrons to flow through the circuit, much like how water flows through a pipe. The velocity at which a light bulb turns on after the flicking of a switch is theoretically the speed of light, but short delays caused by irregularities in the way electrons bump into one another causes the velocity to be roughly 50 to 99 percent of the speed of light.

4.3 Ohm's Law

In the previous section, we developed a microscopic model of charges moving in a conductor, but did not describe how this motion is affected by the electric field in the conductor (or equivalently, the potential difference across the conductor). "Ohm's Law" states that the current density, $\vec{j}$, at some position in the conductor is proportional to the electric field, $\vec{E}$, at that same position in the conductor:

$$\vec{j} \propto \vec{E}$$
$$\boxed{\vec{j} = \sigma \vec{E}}$$

where we have introduced the "conductivity", σ, as the constant of proportionality. Conductivity is a property of the material from which the conductor is made, and is a measure of how large a current density (and by extension, current) there will be in material given a certain electric field. Materials with a high conductivity are said to be good conductors, as a large current will result from a small electric field. Gold and copper are examples of materials with a high conductivity.

Checkpoint 4-1

What is the conductivity of an ideal insulator?

A) 0
B) Roughly 1
C) Infinite

Example 4-3

A gold wire is in an constant electric field $E = 0.5\,\text{NC}^{-1}$. Given that the conductivity of gold is $\sigma = 2.2 \times 10^8\,\Omega^{-1}\text{m}^{-1}$, what is the drift velocity of the electrons in this wire?

Solution

First, we must calculate the current density of the wire:

$$j = \sigma E$$
$$j = 2.2 \times 10^8\,\Omega^{-1}\text{m}^{-1} \cdot 0.5 \times 10^{-6}\,\text{JC}^{-1}$$
$$j = 1.1 \times 10^8\,\text{Am}^{-2}$$

We can now calculate the drift velocity of the electrons in the wire:

$$v_d =$$

4.3.1 Resistivity

For convenience, one often describes how well a material conducts charges using the "resistivity", ρ, which is simply defined as the inverse of conductivity:

$$\rho = \frac{1}{\sigma}$$

Materials with a high resistivity are poor conductors; they tend to "resist" the formation of a current when an electric field is applied. Insulators have high resistivity.

The resistivity of most (but not all) materials has been observed to increase linearly with the temperature of the material. One can picture that as atoms in the material vibrate more, it is more difficult for electrons to conduct through the material as they will interact with more atoms. The resistivity, ρ, at a certain temperature, T, is usually modelled as follows:

$$\rho(T) = \rho_0 \left[1 + \alpha(T - T_0)\right]$$

where, ρ_0, is a "reference resistivity" measured at a "reference temperature", T_0 (usually 20 °C). α is the "temperature coefficient" of the material. The temperature dependence of the resistivity is illustrated in Figure 4.5.

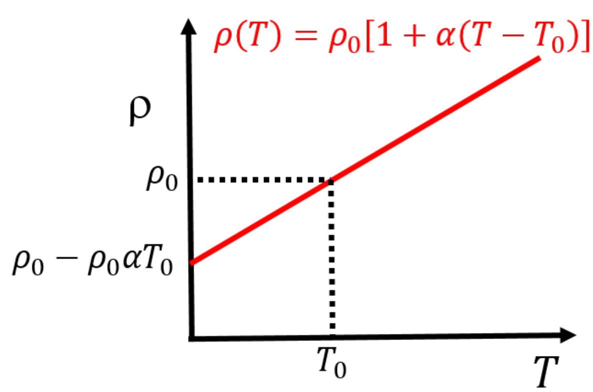

Figure 4.5: A linear model of resistivity can be used for most conductors over a large range of temperatures.

This "linear model" (since resistivity increases linearly with temperature) is empirically found to be valid for many materials over a large range of temperatures, although it is not expected to hold at extreme temperatures (either very low or very high). Furthermore, for semi-conducting materials (such as silicon and germanium), resistivity is found to decrease as a function of temperature.

Checkpoint 4-2

What is the slope of the resistivity vs temperature as shown in Figure 4.5?

A) α

B) $\rho_0 \alpha T$

C) $\rho_0 T$

D) $\rho_0 \alpha$

Table 4.1 shows a list of common materials and their conductivity, resistivity, and temperature coefficients (defined at a reference temperature $T_0 = 20\,^\circ\mathrm{C}$).

Material	**Resistivity** $[\Omega \cdot \mathrm{m}]$	**Temperature coefficient** $[^\circ\mathrm{C}^{-1}]$	**Free electron density** $[\mathrm{m}^{-3}]$
Copper	1.68×10^{-8}	0.0039	8.46×10^{28}
Silver	1.59×10^{-8}	0.0038	5.86×10^{28}
Gold	2.2×10^{-8}	0.0034	5.90×10^{28}
Iron	9.71×10^{-8}	0.0050	17.0×10^{28}
Aluminum	2.74×10^{-8}	0.0039	18.1×10^{28}
Silicon	0.1-60	-0.07	0
Rubber	(1-100) $\times 10^{13}$	0	0
Quartz	7.5×10^{17}	0	0

Table 4.1: Resistivity, free electron density and temperature coefficients of common materials. All properties are listed for a reference temperature of $20^\circ C$

4.4 Resistors

A conductor with current going through it (or current that could go through it) is generally called a "resistor", to emphasize that charges will experience resistance as they travel through the conductor (as they collide with atoms in the resistor). In this section, we describe resistors, how to combine them, and how to model the heat that is generated when charges collide with the atoms in the resistor.

4.4.1 Resistance

Consider a resistor, with length, L, and cross-sectional area, A, made out of a material with resistivity, ρ, as illustrated in Figure 4.6.

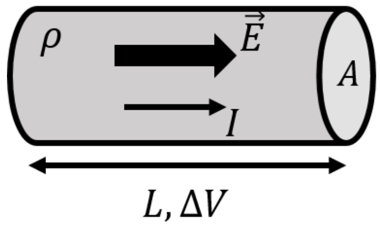

Figure 4.6: A simple resistor of length, L, cross-sectional area, A, made from a materials with resistivity, ρ. A potential difference, ΔV, is applied across the resistor, leading to an electric field and current in the resistor.

A potential difference, ΔV, is applied across the length of the resistor, resulting in an electric field, $\vec{E}$, within its volume. To good approximation, one can model the two ends of the conductor as parallel plates, so that the magnitude of the electric field throughout the conductor is constant in magnitude and direction and has strength given by:

$$E = \frac{\Delta V}{L}$$

Combining this with Ohm's Law, we have:

$$\begin{aligned} j &= \sigma E \\ \therefore j &= \sigma \frac{\Delta V}{L} \end{aligned}$$

Since the current density is a microscopic quantity, we can replace it with the current, I, a macroscopic quantity, for the conductor of cross-sectional area, A, to find:

$$\begin{aligned} j &= \frac{I}{A} \\ \therefore I &= jA = \sigma \frac{\Delta V}{L} A \end{aligned}$$

This last equation is often written by isolating the potential difference:

$$\boxed{\Delta V = \rho \frac{L}{A} I}$$

where we replaced the inverse of the conductivity with the resistivity. This last equation is the equivalent of Ohm's Law, but written for a (macroscopic) resistor of length, L, cross-sectional area, A, and made of a material with resistivity, ρ. Written in this way, Ohm's Law is a statement that the **current through a resistor is proportional to the voltage applied across it**. The constant of proportionality, R, is called the "resistance":

$$\boxed{\Delta V = RI}$$

This last equation is often called "Ohm's Law", even if, technically, Ohm's Law is the relation between current density and electric field. A resistor is a macroscopic object whose "resistance" can be characterized by a single value, R, its resistance. The resistance of a resistor can be determined from its macroscopic properties (length and cross-sectional area) and from the material from which it is made (with a given resistivity):

$$\boxed{R = \rho\frac{L}{A}}$$

The (derived) S.I. unit of resistance is the "Ohm", (Ω).

Checkpoint 4-3

What are the SI units of conductivity?

A) $\frac{\Omega}{\mathrm{C}}$
B) $\frac{1}{\Omega \mathrm{m}}$
C) $\frac{\mathrm{N}^2\Omega}{\mathrm{C}}$
D) $\frac{\mathrm{C}}{\mathrm{s}}$

The model to describe the resistance of a conductor to the flow of electric current under a fixed potential difference, ΔV, is identical to the model that we derived in Section **??** to describe the Poiseuille flow, Q, of an viscous incompressible fluid in a pipe with resistance, R, under a pressure difference, ΔP:

$$\Delta P = RQ$$

Thus, one can think of electric current by analogy to the incompressible flow of a viscous fluid through a pipe. If the pipe is longer, it opposes more resistance to the flow of liquid, just as a longer resistor has a larger resistance to current. A pipe with a larger cross-sectional area has less resistance to the flow of liquid, just as a resistor with a larger cross sectional area, A, has a lower resistance.

4.4.2 Combining resistors

Resistors are the most common component in circuits, and we show below how to model the equivalent resistance of two resistor that are combined in "parallel" or in "series".

Figure 4.7 shows two resistors, R_1 and R_2, connected in "series", to form an an effective resistor with resistance, R_{eff}. A potential difference, ΔV, is applied across the combination of resistors.

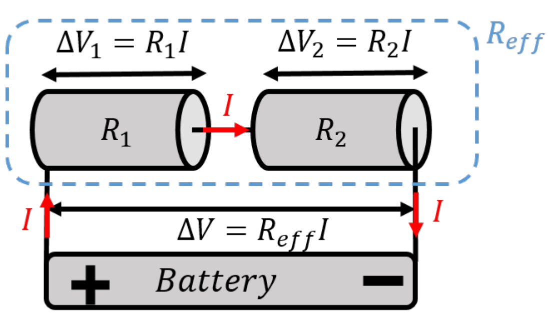

Figure 4.7: When two resistors are connected in series, the same current flows through each resistor.

By analogy with fluid mechanics, the charges that enter resistor, R_1, must exit the resistor at the same rate, and then cross the second resistor, R_2. In other words, what comes into R_1 must come back out of R_2, since there is no place for the charges to go. This is the electrical equivalent of "continuity" in fluid mechanics. **When resistors are combined in series, both resistors will have the same current, I, through them**.

Ohm's Law (the macroscopic version), must also be true for each resistor:

$$\Delta V_1 = R_1 I$$
$$\Delta V_1 = R_2 I$$

where, ΔV_1 and ΔV_2, are the potential differences across each resistor. ΔV_1 and ΔV_2 must sum to ΔV:

$$\Delta V_1 + \Delta V_2 = \Delta V$$

since the potential energy (per unit charge) that is lost in each resistor must equal to the total potential energy (per unit charge) that was lost in going through the combination of resistors. Combining this last equation with Ohm's Law for each resistor, we can model the series combination of resistor as having an "effective resistance", R_{eff}, given by:

$$\Delta V = \Delta V_1 + \Delta V_2 = R_1 I + R_2 I = (R_1 + R_2)I = R_{eff} I$$
$$\boxed{R_{eff} = R_1 + R_2} \quad \text{(Series resistors)}$$

It makes sense that the equivalent resistance if found by summing the two resistors, when these are in series. If the two resistors are made of the same material and have the same cross-sectional area, combining them in series is equivalent to fabricating a longer resistor with the two lengths added together. The result is easily extended to any number of resistors:

$$R_{eff} = R_1 + R_2 + R_2 + \dots$$

Figure 4.8 shows two resistors, with resistances R_1 and R_2, combined in parallel to form an effective resistor with resistance, R_{eff}. A potential difference, ΔV, is applied across the combination of resistors. **When resistors are combined in parallel, both resistors have the same potential difference across them**.

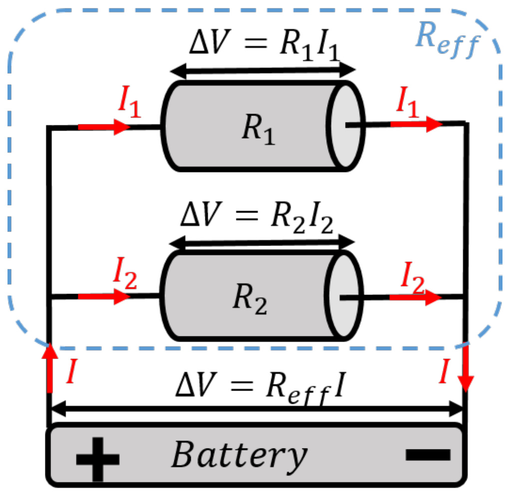

Figure 4.8: When two resistors are connected in parallel, the same voltage is applied across each resistor.

Applying Ohm's Law to each resistor, we find that they each have difference currents going through them:

$$I_1 = \frac{\Delta V}{R_1}$$
$$I_2 = \frac{\Delta V}{R_2}$$

The total current, I, that enters the combination of resistors, must also exit the combination of resistor (continuity), so the total current, I, is the sum of the current through each resistor:

$$I = I_1 + I_2$$

Combining this with Ohm's Law, we find:

$$I = I_1 + I_2 = \frac{\Delta V}{R_1} + \frac{\Delta V}{R_2} = \left(\frac{1}{R_1} + \frac{1}{R_2}\right)\Delta V$$
$$\therefore \Delta V = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2}} I$$

Thus, the effective resistance, R_{eff}, of two resistors connected in parallel is given by:

$$\boxed{R_{eff} = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2}} = \frac{R_1 R_2}{R_1 + R_2}} \quad \text{(Parallel resistors)}$$

where the two forms that are given are equivalent. The effective resistance of two resistors in parallel is smaller than the resistance of either resistor. This makes sense, because combining resistors in parallel is analogous to fabricating a single resistance with a larger cross-sectional area, allowing for "more space" for the charges to flow. Again, this result is easily extended for more than two resistors:

$$R_{eff} = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \dots}$$

Example 4-4

A $R_2 = 2\,\Omega$ resistor is placed in parallel with a $R_3 = 3\,\Omega$ resistor and the combination is placed in series with a $R_1 = 1\,\Omega$ resistor, as shown in Figure 4.9. What is the effective resistance of this combination?

Solution

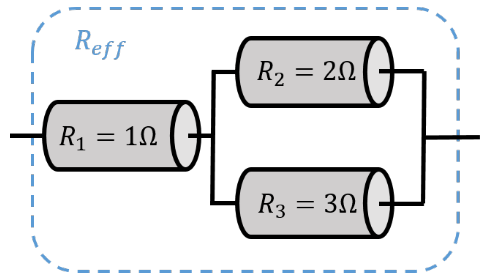

Figure 4.9: A combination of three resistors.

In order to determine the effective resistance of the combination, we can first combine the resistors R_2 and R_3 into an effective resistor, R', which we can then combine in series with the resistor R_1, to obtain the effective resistance of the three resistors. First, combining the parallel resistors, R_2 and R_3, we find:

$$R' = \frac{R_2 R_3}{R_2 + R_3} = \frac{6}{5}\Omega$$

We can then combine this in series with R_1, to obtain the total effective resistance of the combination of three resistors:

$$R_{eff} = R_1 + R' = \frac{11}{5}\Omega$$

Discussion: In this example, we showed how to determine the effective resistance of a combination of series and parallel resistors. We can determine the effective resistance of complex combinations of resistors in the same manner, by first combining subsets of resistors and then including those with other resistors.

4.4.3 Electrical power dissipated in resistors

As we discussed in Section 4.2, charges that move through a resistor do not gain kinetic energy. Instead, the electric potential energy available from the voltage applied across the resistor is converted into heat, as a result of charges colliding with atoms in the material. The net potential energy, ΔU, available to a single charge, q, is given by:

$$\Delta U = q\Delta V$$

If there are many charges going through the resistor, the rate, P, at which they will dissipate energy in the resistor is given by:

$$\begin{aligned} P &= \frac{d}{dt}\Delta U = \frac{d}{dt}q\Delta V = I\Delta V \\ \therefore P &= I\Delta V \end{aligned}$$

where we recognized that $dq/dt = I$. P corresponds to the rate at which energy is dissipated in the resistor, and has dimensions of power. Combining this with Ohm's Law, the power that is dissipated in a resistor can be written in different ways:

$$\boxed{P = I\Delta V = \frac{(\Delta V)^2}{R} = I^2 R}$$

Example 4-5

A hair-dryer is rated as consuming 1500 W when connected to an outlet with a 120 V potential difference. What is the resistance of the hair-dryer, and how much current goes through it when it is running?

Solution

Since the power of the hair-dryer and the potential difference across it are known, we can easily determine its resistance:

$$\begin{aligned} P &= \frac{(\Delta V)^2}{R} \\ \therefore R &= \frac{(\Delta V)^2}{P} = \frac{(120\,\text{V})^2}{(1500\,\text{W})} = 9.6\,\Omega \end{aligned}$$

Similarly, we can determine the current through the hair dryer:

$$\begin{aligned} P &= I\Delta V \\ \therefore I &= \frac{P}{\Delta V} = \frac{(1500\,\text{W})}{(120\,\text{V})} = 12.5\,\text{A} \end{aligned}$$

Discussion: Most household appliances are rated by the electrical power that they consume. This rating assumes that the appliance will be connected to a fixed potential difference (120 V in North America), so it is straightforward to determine the current that they will draw. This is important, because the current that is drawn by the appliance has to go through the wiring in the house, and if the current is too large, the wiring (which has resistance) will heat up ($P = I^2R$) which could result in an electrical fire. Circuits in a house have safety devices (fuses or breakers) that are designed to interrupt the circuit if the current is too large.

One can rate a power supply, such as a battery, by the amount of power that it can deliver. Power supplies are usually designed to supply a fixed potential difference; for example, a 9 V battery supplies a constant voltage of 9 V. If a small resistor is connected across the terminals of the battery, a large current, I, will flow through the resistor. In principle, the current through the resistor will be given by Ohm's Law, $I = \Delta V/R$. However, by making the resistance increasingly smaller, the current will increase, and the power dissipated by the resistor, $P = I\Delta V$, would increase indefinitely. Obviously, this is not possible, as it requires the battery to supply energy at the same ever increasing rate. In practice, as the resistance is decreased, the current through the resistor will only increase until $I\Delta V$ is equal to the maximal power that can be dissipated by the battery. As the resistance across the battery is further decreased, the voltage across the battery will start to decrease as well, so that the power dissipated in the resistor, $\Delta V I$, does not exceed the power that the battery could possibly supply.

4.4.4 Superconductors

Superconductors are materials that, under certain conditions, have zero resistivity. A resistor made from a superconducting material will thus have zero resistance. It is beyond the scope of this textbook to describe how superconductivity arises in materials, however, it is worth knowing that these exist. Typically, superconductivity arises in materials when they are cooled to temperatures close to absolute zero, although some materials exhibit superconductivity at much higher temperatures ($\sim 140\,^{\circ}\mathrm{K}$ ($\sim -130\,^{\circ}\mathrm{C}$). Superconducting materials are often used when one needs a large electric current, such as in a powerful electro-magnet. By having no resistance, a large current can be sustained without dissipating any power.

4.5 Alternating voltages and currents

So far, we have modelled how current propagates through a resistor under a constant potential difference, ΔV. This is called "direct current" (DC) as the charges move in a constant direction through the resistor. Batteries supply fixed voltages, and circuits with batteries will almost always have DC current. The voltage that is supplied between two of the sockets in a household electrical outlet is "alternating", and leads to "alternating current" (AC), where charges move back and forth, with no net displacement.

The potential difference across a household outlet varies sinusoidally:

$$\Delta V(t) = \Delta V_0 \sin(\omega t)$$

where ΔV_0 is the maximal amplitude of the voltage (120 V in North America, 220 V in Europe), and $\omega = 2\pi f$, is the angular frequency of the voltage ($f = 60\,\mathrm{Hz}$ in North America, $f = 50\,\mathrm{Hz}$ in Europe). When a resistor with resistance, R, is connected to an AC voltage, the resulting current, given by Ohm's Law, is also alternating:

$$I(t) = \frac{\Delta V(t)}{R} = \frac{\Delta V_0}{R}\sin(\omega t) = I_0 \sin(\omega t)$$

On average, the alternating current through a resistor is zero. However, this does not mean that zero energy is dissipated, since the electrons in the resistor will still collide with atoms as they oscillate back and forth. We can define the average power, $\bar{P}$, that is dissipated in the resistor as the power that is dissipated over one oscillation cycle (with period, T). To obtain the latter, we calculate the total energy, E, dissipated in the resistor over one cycle so that the power is simply given by E/T. We divide the interval of time, T, into infinitesimally small intervals, dt, so that the infinitesimal energy, dE, dissipated in an infinitesimal time, dt, is given by:

$$dE = P(t)dt$$

The total energy dissipated in one period is then given by:

$$E = \int dE = \int_0^T P(t)dt$$

so that the power dissipated in one cycle is given by:

$$\bar{P} = \frac{E}{T} = \frac{1}{T}\int_0^T P(t)dt$$

The instantaneous power, $P(t)$, can be described in terms of the instantaneous current, $P(t) = I^2(t)R$, so that the average power can be written as:

$$\bar{P} = \frac{1}{T}\int_0^T P(t)dt = \frac{1}{T}\int_0^T I(t)^2 R dt = RI_0^2 \frac{1}{T}\int_0^T \sin^2(\omega t)dt = \frac{1}{2}RI_0^2$$

where we used the fact that $T = \frac{2\pi}{\omega}$ to evaluate the integral. In order to make the formula similar to the DC equivalent (without the additional factor of 1/2), we can define the "root mean square" current, I_{rms}, as an average current, from which we can calculate the average power that is dissipated in a resistor:

$$\begin{aligned} I_{rms} &= \frac{I_0}{\sqrt{2}} \\ \therefore \bar{P} &= I_{rms}^2 R \end{aligned}$$

Similarly, one can define the "root mean square" voltage, ΔV_{rms}, so that the average power dissipated with alternating current can be written in the same form as for the DC case:

$$\begin{aligned} V_{rms} &= \frac{\Delta V_0}{\sqrt{2}} \\ \therefore \bar{P} &= I_{rms}^2 R = \frac{\Delta V_{rms}^2}{R} = I_{rms}\Delta V_{rms} \end{aligned}$$

4.6 Electrical safety

The models that we have developed to describe current can inform us on ways to avoid being injured by electricity in our common lives. The two main hazards associated with electricity are fire and electrocution. Typically, an electrical fire is the result of a large current going through a resistor, as the power dissipated in a resistor is proportional the square of the current through that resistor. If you connect an appliance that draws a large current to your outlets, the wires in your house could heat up enough to cause a fire. This danger is primarily mitigated by using "fuses" or "circuit breakers" that will interrupt the circuit if the current is too large. A fuse is a simple device with a thin wire (high resistance) that will melt and break if too much current goes through it (which is designed to happen long before the wires in your house start to overheat). A circuit breaker is a resettable switch that opens under a large current. Modern houses do not use fuses any more, since they have to be replaced every time they are "blown".

Electrocution is a form of injury that is the result of a current crossing the body; we can think of the body as a resistor connected between the terminals of a battery. Injuries can be caused simply by burns (tissue destroyed), or by muscles contracting involuntarily due to the current. For example, one's muscles may contract in such a way that the person cannot let go of the source of current. If a current of more than about $80\,\text{mA}$ passes through the mid section of a person, enough current could go through the heart so that it starts to beat very irregularly ("ventricular fibrillation") which can lead to death since blood stops flowing normally. A very large current can cause the heart to simply stop beating, which could sometimes be less dangerous than ventricular fibrillation (if for a short period of time, and of course, the burns will be more severe from a larger current). A "defibrillator" is designed to provide such a high current that the heart stops briefly, with the hope that when it starts back, the beats will be regular. This can be used in cases of ventricular fibrillation. One often hears that "it's current that kills", which is a statement that being electrocuted by a certain voltage is not a good measure of the resulting injury, since the current will depend on the resistance of the person's body.

The amount of current that will go through a person will depend on the resistance of the person's body. Internal tissues and organs are typically quite conductive and have low resistance. The outer layer of the skin, on the other hand, has a high resistance when dry and helps to limit the current that can go through the body. The resistance of dry skin is usually considerably above $1 \times 10^4\,\Omega$, while it can be much less than $1 \times 10^3\,\Omega$ when wet. With wet skin, a potential difference of $120\,\text{V}$ (as in a North American outlet) can easily lead to a current above $100\,\text{mA}$, which could easily be fatal. Note that being barefoot and in contact with the ground is usually a low resistance connection, since there is often a thin layer of sweat on your feet.

In North America, electrical outlets have a minimum of two sockets: a "live" socket (with an oscillating voltage, usually a black wire[1]), and a "neutral" socket which is connected to the ground and relative to which the oscillating voltage has an amplitude of $12\,\text{V}$ (usually a white wire). One can obviously be electrocuted by simultaneously touching the wires in both sockets, and usually simply by touching the wire in the live socket, since one's feet are usually connected to ground. Electrocution by directly touching the socket is fairly uncommon, since most people know not to do that (right?!). Usually, one is electrocuted by an appliance with faulty wiring; perhaps the insulation on the live wire is worn out and you touch the wire by mistake, or the wiring in the appliance is faulty, causing the casing of the appliance to be live. In order to mitigate the risk of electrocution from an appliance with

[1] Never trust the colouring of wires, always test them!

faulty wiring, most outlets will have a third socket, the "dedicated ground". The dedicated ground wire is connected to the ground inside the socket, and to the casing of the appliance, as illustrated in Figure 4.10. Thus, if the live wire were to be in contact with the casing of the appliance, the dedicated ground provides a low resistance path for current to take that is in parallel with your body (so that most current will go through the low resistance path).

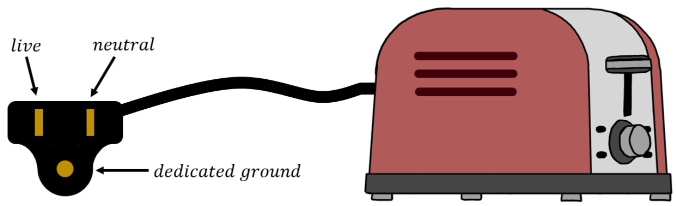

Figure 4.10: When an appliance has three prongs on its electrical cable, the middle prong grounds the case to the dedicated ground as a safety measure. Note that the live wire is not necessarily the left socket on an outlet!

4.7 Summary

Key Takeaways

Electric current, I, is defined as the rate at which charges cross some plane (for example a plane perpendicular to a wire) per unit time. That is, if an amount of charge, ΔQ, enters a wire during an amount of time, Δt, the current, I, in that wire is defined to be:

$$I = \frac{\Delta Q}{\Delta t} = \frac{dQ}{dt}$$

where a derivative is taken if the rate at which charges are moving is not constant with time.

Electric current is a macroscopic quantity that can be measured. Conventional current is defined to be positive in the direction in which positive charges flow. In most situations, it is electrons that move inside a conductor, so the current is defined to be positive in the *opposite direction* of the actual motion of the (negative) electrons.

The current density, $\vec{j}$, is defined to be the current per unit area at some point in a conductor, and is a vector in the direction of the electric field, $\hat{E}$, at that point:

$$\vec{j} = \frac{I}{A}\hat{E}$$

The current density can be related to the microscopic motion of charges within the conductor. If the current density, $\vec{j}$, is known, the corresponding current, I, that crosses a surface with area, A, and normal vector, $\hat{n}$, is given by:

$$I = A\vec{j} \cdot \hat{n} = \int \vec{j} \cdot d\vec{A}$$

where the integral must be taken if the current density is not constant over the surface.

A conducting material through which current is flowing is called a resistor. When a potential difference is applied across a resistor, the resulting electric field will drive the flow of electrons through the resistor. The electrons will flow with an average "drift velocity", $\vec{v}_d$, which is much slower than the actual motion of the electrons. Inside the resistor, electrons are constantly accelerated before they collide with atoms in the material losing their kinetic energy, and then accelerating again. Thus, the potential energy that is available to the electrons is "used" to heat the resistor, and the electrons, on average, drift quite slowly through the resistor.

The current density in a resistor can be related to the drift velocity of the electrons and the "density of free electrons" in the material, n:

$$\vec{j} = -en\vec{v}_d$$

where, e, is the magnitude of the charge of the electrons and the minus sign indicates that the current density is in the opposite direction of the velocity of the (negative) electrons.

Ohm's Law states that the current density, $\vec{j}$, at some point in the conductor is proportional to the electric field, $\vec{E}$, at that point:

$$\vec{j} = \sigma\vec{E} = \frac{1}{\rho}\vec{E}$$

where the constant of proportionality, σ, is called the "conductivity" of the material (and is a property of the material through which current is flowing). The resistivity, ρ, is a material property that is simply the inverse of the conductivity. Both of these properties are a measure of how large a current (or current density) will exist in a material given a certain electric field. For example, the conductivity of an insulating material is close to zero (and its resistivity close to infinity).

For most materials, resistivity usually increases linearly with temperature:

$$\rho(T) = \rho_0[1 + \alpha(T - T_0)]$$

where ρ_0 is the resistivity as measured at some reference temperature, T_0 (usually 20 °C), and α, is the "temperature coefficient" for that material. Note that this model of resistivity does not hold for extreme temperatures (very cold or very hot), and for some materials, resistivity decreases with temperature (α is negative).

If we apply Ohm's Law to a resistor of length, L, cross-sectional area, A, made of a material with resistivity, ρ, we find that the potential difference applied across the resistor, ΔV, is proportional to the current flowing through the resistor:

$$\Delta V = \rho\frac{L}{A}I$$

The constant of proportionality depends on the material with which the resistor is made (through the resistivity) and on the dimensions of the resistor (through the length and cross-sectional area). The constant of proportionality is called the "resistance" of the resistor, R:

$$R = \rho\frac{L}{A}$$

Ohm's Law is often written for a resistor as the relationship between the current through the resistor, I, and the potential difference across the resistor, ΔV:

$$\Delta V = RI$$

although, technically, Ohm's Law is the relation between current density and electric field.

Resistors can be combined in series, in which case, the effective resistance of the combination is found by adding the resistances of the individual resistors:

$$R_{eff} = R_1 + R_2 + R_2 + \dots \quad \text{(Series resistors)}$$

When combined in parallel, the inverse of the effective resistance is given by the inverse of the sum of the inverse of the resistances of the individual resistors:

$$R_{eff} = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \dots} \quad \text{(Parallel resistors)}$$

As charges move through a resistor of resistance, R, under a potential difference, ΔV, and current, I, they transfer their kinetic energy into heating up the resistor. The rate at which they transfer the energy, also called the "power dissipated in the resistor", is given by:

$$P = I\Delta V = \frac{(\Delta V)^2}{R} = I^2 R$$

where the various combinations can be obtained by applying the macroscopic version of Ohm's Law.

The electrical outlets in our daily lives provide an "alternating" voltage, $\Delta V(t)$, which oscillates sinusoidally:

$$\Delta V(t) = \Delta V_0 \sin(\omega t)$$

with a maximum amplitude, ΔV_0, and an angular frequency, $\omega = 2\pi f$. When this potential difference is applied across a resistor, an alternating current is formed, in which the electrons move back and forth, with no net displacement:

$$I(t) = \frac{\Delta V_0}{R} = I_0 \sin(\omega t)$$

Even though there is not net displacement, the electrons will still transfer energy into the resistor in the form of heat. The average rate at which power is dissipated in the resistor is given by:

$$\bar{P} = \frac{1}{2} R I_0^2$$

We introduce the "root mean square" current (voltage), I_{rms} (ΔV_{rms}), as an average effective current (voltage):

$$I_{rms} = \frac{1}{\sqrt{2}} I_0$$
$$\Delta V_{rms} = \frac{1}{\sqrt{2}} \Delta V_0$$

such that the power can be expressed using a similar formula as in the direct current case, using the root mean square values:

$$\bar{P} = I_{rms}^2 R = I_{rms} \Delta V_{rms} = \frac{(\Delta V_{Rms})^2}{R}$$

There are two main types of hazards associated with the use of electricity: fire and electrocution. Electrical fires can arise when a large current goes through a wire, since this will dissipate a large amount of heat into the wire (which could set fire to its insulation or other nearby flammable items). Electrocution occurs when a current traverses the human body. If a current above $\sim 80\,\mathrm{mA}$ crosses the upper body, this can result in ventricular fibrillation, whereby the heart beats very irregularly. Of course, one can also be burned by a large current. The amount of current through the body is what will ultimately determine the severity of injuries, and is why one often hears that "it's current that kills". A large voltage may not lead to a large current if the resistance of the person is large or if the power supply cannot provide a large current at that large voltage.

Important Equations

Current:

$$I = \frac{\Delta Q}{\Delta t}$$

$$I = \int \vec{j} \cdot d\vec{A}$$

Voltage:

$$\Delta V = RI$$

Current density:

$$\vec{j} = -en\vec{v}_d$$

Resistance:

$$R = \rho \frac{L}{A}$$

Resistivity:

$$\rho(T) = \rho_0[1 + \alpha(T - T_0)]$$

$$\rho = \frac{1}{\sigma}$$

Important Definitions

Current: The rate at which charge flows through a two dimensional region. SI units: [A]. Common variable(s): I.

Current density: A measure of current per unit area. SI units: [Am^{-1}]. Common variable(s): $\vec{j}$.

Resistance: A measure of a specific object's opposition to the flow of charge. SI units: [Ω]. Common variable(s): R.

Resistivity: A measure of a material's opposition to the flow of charge. SI units: [Ωm]. Common variable(s): ρ.

Conductivity: The inverse of resistivity. SI units: [$\Omega^{-1}\mathrm{m}^{-1}$]. Common variable(s): σ.

Drift velocity: The average net velocity of an electron in a conductor. SI units: [ms^{-1}]. Common variable(s): $\vec{v}_d$.

4.8 Thinking about the material

Reflect and research

1. Describe how superconductivity arises in certain materials (hint: research "Cooper pairs").
2. What are some example of superconducting materials, and at what temperature do they become superconducting?
3. Is there a limit to how much current a conductor can carry?
4. Does an AC circuit have a draft velocity? Why or why not?

To try at home

1. Use Ohm's law and the electrical information on an appliance to measure the current produced by your home's electrical outlets.
2. What is the current produced by your phone's battery? What is the total power stored in your phone's battery? Check the technical information to compare your results to the manufacturer's measurements.
3. Try

To try in the lab

1. Propose an experiment to create an AC circuit and measure its current.
2. Propose an experiment to measure the α and resitivity of various wires.

4.9 Sample problems and solutions

4.9.1 Problems

Problem 4-1: (Solution) A special type of cylindrical wire is constructed such that its current density increases at a rate of $j = 900r$, where r is measured from the centre of the wire's circular cross-section. If the wire has a radius of $r = 1.5\,\text{cm}$, what is the current in the wire?

Problem 4-2: A resistor has a resistivity of $1.65 \times 10^{-8}\,\Omega\text{m}$ at a temperature of $30\,°\text{C}$ and has a resistivity of $1.53 \times 10^{-8}\,\Omega\text{m}$ at a temperature of $10\,°\text{C}$. What material is this resistor made of? (Solution)

4.9.2 Solutions

Solution to problem 4-1: To solve for the current, we must integrate over the area of the wire's cross section:

$$
\begin{aligned}
I &= \int j dA \\
I &= \int j r dr d\theta \\
I &= \int 900 r^2 dr d\theta \\
I &= \int_0^{2\pi} \int_0^{0.015} 900 r^2 dr d\theta \\
I &= \int_0^{2\pi} 900 \cdot 1.125 \times 10^{-6} d\theta \\
I &= 2\pi \cdot 900 \cdot 1.125 \times 10^{-6} \\
I &= 6.36 \times 10^{-3}\,\mathrm{A}
\end{aligned}
$$

Which gives a final answer of 6.36 mA.

Solution to problem 4-2: To solve for the material of the resistor, we must find the temperature coefficient, α. We can solve for this because we are given two measurements of resistivity at different temperatures. The reference temperature and resistivity measurements are arbitrarily chosen to be $1.53 \times 10^{-8}\,\Omega\mathrm{m}$ at 10 °C.

$$
\begin{aligned}
\rho(T) &= \rho_0[1 + \alpha(T - T_0)] \\
\frac{\frac{\rho(T)}{\rho_0} - 1}{T - T_0} &= \alpha \\
\frac{\frac{1.65 \times 10^{-8}\,\Omega\mathrm{m}}{1.53 \times 10^{-8}\,\Omega\mathrm{m}} - 1}{30\,^\circ\mathrm{C} - 10\,^\circ\mathrm{C}} &= \alpha \\
\alpha &= 0.0039\,^\circ\mathrm{C}^{-1}
\end{aligned}
$$

Now that we have α, we can calculate the resistivity of the material at 20 °C, which we can then compare to the list of materials and their resistivity:

$$
\begin{aligned}
\rho(T) &= \rho_0[1 + \alpha(T - T_0)] \\
\rho(20) &= 1.53 \times 10^{-8}\,\Omega\mathrm{m}[1 + 0.0039\,^\circ\mathrm{C}^{-1}(20\,^\circ\mathrm{C} - 10\,^\circ\mathrm{C})] \\
\rho(20) &= 1.59 \times 10^{-8}\,\Omega\mathrm{m}
\end{aligned}
$$

Now that we have found the resistivity of the mystery material, we can refer to the list of materials and their resistivities. We will find that silver also has a resisitivity of $1.59 \times 10^{-8}\,\Omega\mathrm{m}$ at 20 °C, which means that the unknown material in the resistor is silver.

5 Electric circuits

In this chapter, we develop the tools to model electric circuits. This will allow us to determine the current and voltages across different components, resistors and capacitors, within a circuit. We will also discuss how a battery can provide a current at a fixed potential difference, and how one can construct devices to measure current and voltages.

Learning Objectives

- Understand how a battery works.
- Understand Kirchhoff rules and how to apply them.
- Understand how to model a circuit with resistors and/or capacitors.
- Understand how an ammeter and voltmeter function, and how to model them.

Think About It

If two outlets in your house are connected to the same circuit, are the outlets connected in series or in parallel?

A) series
B) parallel

5.1 Batteries and simple circuits

A battery is an electric component that provides a constant electric potential difference (a fixed voltage) across its terminals. Luigi Galvani was the first to realize that certain combination of metals placed into contact with each other can lead to an electric potential difference (or rather, they can make the legs of a dead frog twitch, which we now understand to be from the potential difference due to the metal). Effectively, Galvani created the first "electrochemical cell". Alessandro Volta then combined several of these cells together to form the "voltaic pile", which is what we would now call a battery (a battery, technically, is a combination of several cells, a battery of cells, although one often uses the term battery even if only a single electric cell is involved).

5.1.1 The electrochemical cell

An electric cell can be constructed from metals that have different affinities to be dissolved in acid. A simple cell, similar to that originally made by Volta, can be made using zinc and carbon as the "electrodes" (Volta used silver instead of carbon) and a solution of dilute sulfuric acid (the liquid is called the "electrolyte"), as illustrated in Figure 5.1. Before the cell is constructed, the electrodes and the electrolyte are all electrically neutral.

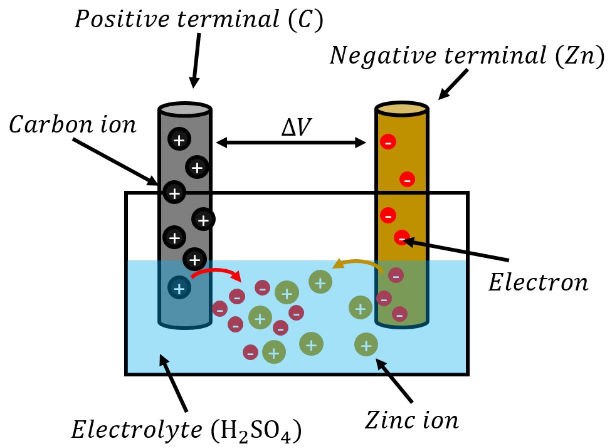

Figure 5.1: A simple electric cell, where zinc ions dissolve in sulfuric acid leaving electrons on the metal.

Once the zinc is immersed in the electrolyte, the zinc atoms tend to dissolve into the electrolyte in the form of zinc ions (doubly charged, Zn^{2+}). This leaves an excess of electrons on the zinc electrode, resulting in a net negative electric charge. Similarly, the positively charged zinc ions attract electrons from the carbon electrode into the solution, leaving the carbon electrode positively charged. Very quickly, an equilibrium is reached, since at some point, the negative charge of the zinc electrode will electrically attract positive zinc ions, preventing any more zinc ions from dissolving into the solution. Similarly, as the carbon electrode builds a positive charge, that charge will eventually prevent electrons from "jumping" into the solution. At this point, there will be a fixed electric potential difference between the two electrodes (terminals) of the battery.

If the two electrodes are connected together through a resistor, the electrons will leave the zinc electrode, cross the resistor, and end up on the positive carbon electrode. This will leave space for more electrons on the zinc electrode, so more zinc ions will dissolve into the solution. Thus, a circuit is formed, where electron travel up the zinc electrode, through the resistor and back down the carbon electrode. At the same time, more and more zinc ions dissolve into the electrolyte, until the zinc electrode is completely dissolved. In practice, the zinc ions travel through the solution and plate onto the carbon electrode (the electrons do not quite "jump" into the electrolyte, rather, it is the zinc ions that move in the electrolyte). Since the charge on the electrodes is continuously replenished, the potential difference between the electrodes remains constant even as current is flowing.

The electric cell will stop working once the zinc electrode has completely dissolved (this is what happens when your battery is dead). Note that there is also a maximum current that the cell can supply, which depends on the rate at which the zinc can dissolve into the electrolyte and plate onto the carbon electrode. If the electrodes of the cell are connected with a very low resistance resistor, the resulting current will be too large for the potential difference to be maintained. Most electric cells work in similar ways, although the chemical reactions can be much more complex. Sometimes, the chemical reaction is reversible; one could use a different battery to apply a negative voltage to the carbon electrode to reverse the reaction and plate the zinc back onto the zinc electrode, thus "recharging the battery" (and converting electric energy back into stored chemical potential energy).

5.1.2 The ideal battery in a circuit

As we proceed, we will use the term "battery" loosely to refer to a device (such as an electric cell or collection of cells) that can provide a fixed potential difference between two terminals (or electrodes). Figure 5.2 shows the circuit diagram for a battery, consisting in two (or four) vertical bars, with the larger bar indicating the positive terminal of the battery.

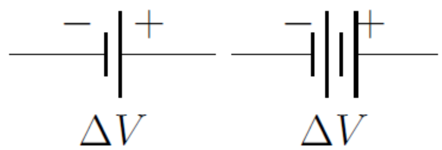

Figure 5.2: Circuit diagram symbols that can be used for a battery.

Figure 5.3 shows the circuit diagram symbols that are used for a resistor (different symbols are used in North American and in Europe).

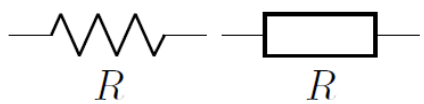

Figure 5.3: Circuit diagram symbols for a resistor, using the North American convention (left), and the European convention (right).

Figure 5.4 shows a circuit diagram for a very simple circuit consisting of a single 9 V battery connected to a 2 Ω resistor. When drawing a circuit diagram (or making a real circuit), one connects the various components together (e.g. batteries and resistors) with **segments of wire that have zero resistance**, even if, in practice, wires always have some resistance. However, since the wires are connected in series with resistors (or other components that have a resistance), one can always include the resistance of the wires by adding it to the resistance of the other components. For example, in Figure 5.4, if the wires have a total resistance of 1 Ω, we could simply model the circuit as if the resistor had a resistance of 3 Ω instead of 2 Ω. In practice, this is usually accounted for when a circuit diagram is made (i.e. any resistors include the resistance of the wires connected to it).

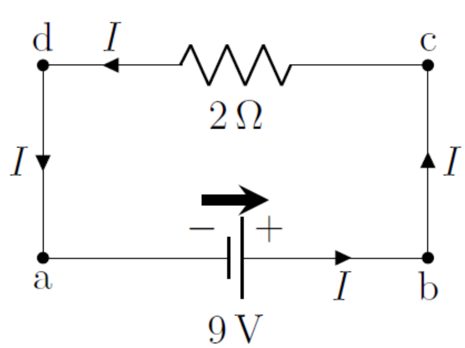

Figure 5.4: A simple circuit, showing a $9\,\mathrm{V}$ battery and a $2\,\Omega$ resistor. For ease in analyzing circuits, we suggest drawing a "battery arrow" above batteries that goes from the negative to the positive terminal.

The circuit in Figure 5.4 is simple to analyze. In this case, whichever charges exit one terminal of the battery, must pass through the resistor and then enter the other terminal of the battery. We **always use conventional current** to analyze a circuit. Thus, we model the circuit as if positive charges exit the positive terminal of the battery, go through the resistor, and then enter the negative terminal of the battery.

We recommend that you always draw a "battery arrow" for each battery in a circuit diagram to indicate the direction in which the electric potential increases and in which conventional current would exit the battery if a simple resistor were connected across the battery. In complex circuits, the current may not necessarily flow in the same direction as the battery arrow, and the battery arrow makes it easier to analyze those circuits. We also indicate the current that is flowing in any wire of the circuit by drawing an arrow in the direction of current on that wire (labelled I in Figure 5.4).

It is helpful to think of the value of the electric potential along different parts of a circuit, as illustrated in Figure 5.5 for the same circuit as in Figure 5.4.

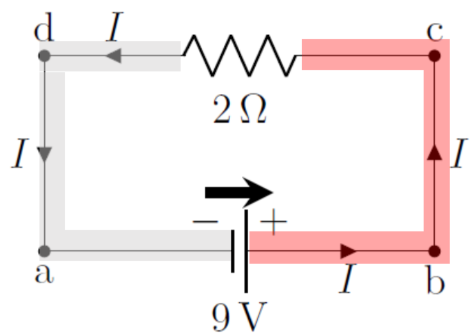

Figure 5.5: The same circuit as in Figure 5.4 showing the two regions over which the electric potential is constant.

Since the wires have no resistance, the electric potential is constant along a wire. In other words, because the wire has no resistance, the charges/current cannot dissipate any power in the wire ($P = I^2R$), and the charges do not "loose" any potential energy (and the potential thus cannot change). The only place where the charges can dissipate energy is inside the resistor. Once the charges have crossed the resistor, the electric potential in the wire is again constant until they reach the other terminal of the battery. Thus, in this simple circuit, the electric potential difference across the resistor is the same as the potential difference across the terminals of the battery. This is shown by the coloured areas in Figure 5.5. If we choose $0\,\mathrm{V}$ to be defined at the negative terminal of the battery, then the potential is $9\,\mathrm{V}$ everywhere in the red area (to the right of the resistor), and $0\,\mathrm{V}$ everywhere in the grey area (to the left of the resistor).

We can apply Ohm's Law (the macroscopic version) to the resistor and determine the current in the circuit, since we know the potential difference across the resistor:

$$\Delta V = RI$$
$$\therefore I = \frac{\Delta V}{R} = \frac{(9\,\text{V})}{(2\,\Omega)} = 4.5\,\text{A}$$

It is helpful to think of circuits in terms of energy. Charges move along the circuit and their potential energy changes as they go through components, while it remains constant as they move through a wire. If a positive charges enters the negative terminal of a battery and exits the positive terminal, its potential energy will have increased. If that charges then enters a resistor, its potential energy will have decreased when exiting the resistor, since the charge will have used its potential energy to heat up the resistor. Batteries provide the energy to "push" the charges through the resistors in the circuit by converting chemical potential energy into the electrical potential energy of the charges.

It is also useful to make the analogy with fluid dynamics; one can think of the battery as a pump that is continuously pushing a viscous incompressible fluid through a pipe with a narrow section, as illustrated in Figure 5.6. The wide section of the pipe is akin to the wires with no resistance, and the narrow section is akin to the resistor. The pressure difference generated by the pump is analogous to the voltage produced by the battery, and the flow rate of the liquid is analogous to the electric current. The pressure in the pipe does not drop in the wide section, if there is no resistance. The entire pressure drop of the fluid is across the narrow section, just as the voltage only drops across the resistor.

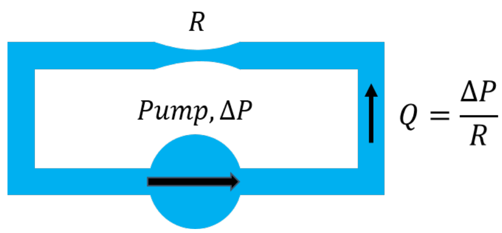

Figure 5.6: A fluid dynamics analogue of the circuit in Figure 5.4, where a pump plays the role of the battery, and a narrow pipe that of a resistor.

Example 5-1

Two resistors, of $2\,\Omega$ and $4\,\Omega$, respectively, are connected in series to a $12\,\text{V}$ battery. What is the current through each of the resistors, and what is the voltage across each resistor?

Solution

We start by making a circuit diagram, as in Figure 5.7, showing the resistors, the current, I, the battery and the battery arrow. Note that since this is a closed circuit with only one path, the current through the battery, I, is the same as the current through the two resistors.

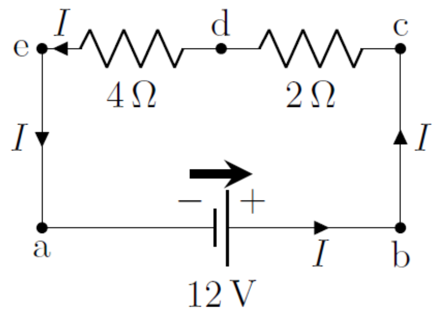

Figure 5.7: Two resistors connected in series with a battery.

If we choose the potential on the negative side of the battery to be 0 V, then points a and e on the diagram are at a potential of 0 V, since potential cannot change in a wire with no resistance. Similarly, the points at b and c are at a potential of 12 V (relative to points a and e). At point d, between the two resistors, the potential will be between 0 V and 12 V, since the potential will "drop" as the current goes through the 2 Ω resistor.

The easiest way to determine the current through this simple circuit is to combine the two resistors into a single effective resistor with resistance:

$$R_{eff} = (2\,\Omega) + (4\,\Omega) = 6\,\Omega$$

so that the circuit can be simplified to that shown in Figure 5.8:

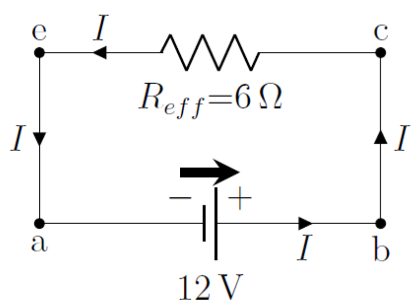

Figure 5.8: The resistors from the circuit in Figure 5.7 have been combined in series to simplify the circuit.

The potential difference across the effective resistor is the same as that across the battery (between points e and c), so that Ohm's Law can be applied to the effective resistor to determine the current that traverses it:

$$\Delta V = R_{eff} I$$
$$\therefore I = \frac{\Delta V}{R_{eff}} = \frac{(12\,\mathrm{V})}{(6\,\Omega)} = 2\,\mathrm{A}$$

This current is the same that traverses each individual resistor, since it is the same as the current that goes through the battery. Referring back to the full circuit (Figure

5.7), we can now use Ohm's Law to calculate the voltage drop across each resistor, since we know the current through each resistor. The voltage across the $2\,\Omega$ resistor is given by:

$$\Delta V_{2\Omega} = RI = (2\,\Omega)(2\,\mathrm{A}) = 4\,\mathrm{V}$$

and the voltage across the $4\,\Omega$ resistor is given by:

$$\Delta V_{4\Omega} = RI = (4\,\Omega)(2\,\mathrm{A}) = 8\,\mathrm{V}$$

Note that the sum of these two voltages is equal to the voltage increase across the battery, by conservation of energy. Consider the electric potential at different points in Figure 5.7 as you move clockwise around the loop starting at point a. If the electric potential is defined to be $0\,\mathrm{V}$ at the negative end of the battery (points a and e), the potential at point d (between the resistors) is the potential at point e plus the potential difference across the $4\,\Omega$ resistor:

$$V_d = V_e + \Delta V_{4\Omega} = (0\,\mathrm{V}) + (\Delta V_{4\Omega}) = 8\,\mathrm{V}$$

If we then add the potential difference across the $2\,\Omega$ resistor to the potential at point d, we find that the potential at point c is $V_c = V_d + \Delta V_{2\Omega} = 12\,\mathrm{V}$, as expected, since this corresponds to the potential at the positive terminal of the battery.

Discussion: In this example, we showed how one can model a circuit by combining resistors together into effective resistors to simplify the circuit. We also showed how the potential differences across different components in a circuit must add up to zero (the voltage drops across the resistors must sum to the voltage increase across the battery).

Checkpoint 5-1

What is the voltage across a $3\,\mathrm{V}$ battery connect in series with a $6\,\mathrm{V}$ battery if the negative terminal of the $6\,\mathrm{V}$ battery faces the positive terminal of the $3\,\mathrm{V}$ battery?

A) $9\,\mathrm{V}$.
B) $6\,\mathrm{V}$.
C) $3\,\mathrm{V}$.
D) $0\,\mathrm{V}$.

5.1.3 The real battery in a circuit

So far, we have modelled batteries as "ideal" devices that provide a fixed potential difference. In reality, this neglects the fact that the materials that make the battery will themselves have a resistance. For example, if electrons want to leave the zinc rod in the electric cell illustrated in Figure 5.1, they will loose some energy as they pass through the zinc. Thus, when modelling a real battery in a circuit, it is important to include its "internal resistance", as a resistor in series with the potential difference. This is illustrated in Figure 5.9, which shows the two terminals of a real battery, an ideal battery (with a fixed potential difference,

ΔV_{ideal}), and its internal resistance, r (which can be drawn on either side of the battery).

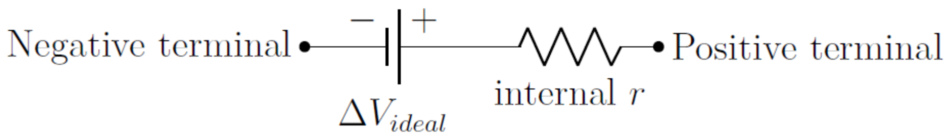

Figure 5.9: Circuit diagram symbol for a battery.

It is important to note that the potential difference across the terminals of the real battery is only equal to the potential difference across the ideal battery **if there is no current flowing through the battery**. If there is a current, I, flowing through the internal resistance, the electric potential will decrease by an amount Ir across the internal resistance, and the voltage across the real terminals will no longer be the same as ΔV_{ideal}.

Example 5-2

When no resistance is connected across a real battery, the potential difference across its terminals is measured to be 6 V. When a $R = 2\,\Omega$ resistor is connected across the battery, a current of 2 A is measured through the resistor. What is the internal resistance, r, of the battery, and what is the voltage across its terminals when the $R = 2\,\Omega$ resistor is connected?

Solution

The real battery can be modelled as an ideal battery with potential difference, ΔV_{ideal}, in series with an internal resistance, r. While we do not know the value of the internal resistance, we are told that the potential difference across the terminals of real battery is 6 V **when no current flows through it**. Since no current flows through the internal resistance, the voltage does not drop across the internal resistance, and the voltage across the terminals of the real battery (e.g. Figure 5.9) must thus be equal to the voltage across the terminals of the ideal battery, so that $\Delta V_{ideal} = 6\,\text{V}$.

With this information, we can make a circuit diagram for the case when the $2\,\Omega$ resistor is connected across the terminals of the real battery, as in Figure 5.10.

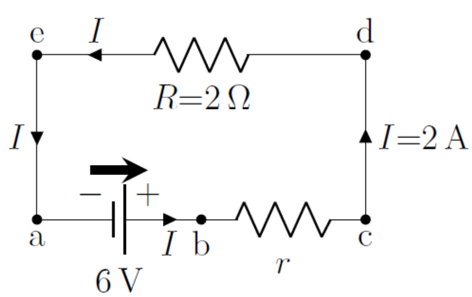

Figure 5.10: A circuit showing a real battery (with internal resistance r) in series with a resistor.

The terminals of the real battery are located at points a and c of the diagram, whereas the terminals of the ideal battery corresponds to points a and b. When no current flows through the internal resistor, there is no voltage drop across that resistor and the potential at b will be equal to the potential at c, as we argued above.

The circuit in Figure 5.10 is now identical to that analyzed in Example 5-1, and can be treated the same way. We can combine the $2\,\Omega$ resistor with the internal resistance, r, in series to obtain an effective resistor, $R_{eff} = r + R$. The voltage drop across the effective resistor will be the same as the potential difference across the ideal battery, and we can make use of Ohm's Law to find the internal resistance, r:

$$\begin{aligned}\Delta V_{ideal} &= R_{eff} I = (r + R)I \\ \therefore r &= \frac{\Delta V_{ideal}}{I} - R = \frac{(6\,\mathrm{V})}{(2\,\mathrm{A})} - (2\,\Omega) = 1\,\Omega\end{aligned}$$

Now that we know the internal resistance, we can determine the voltage drop across the internal resistor, using Ohm's Law:

$$\Delta V_r = rI = (1\,\Omega)(2\,\mathrm{A}) = 2\,\mathrm{V}$$

The voltage drop across the real terminals of the battery (between points a and c), is thus given by:

$$\Delta V_{real} = \Delta V_{ideal} - \Delta V_r = (6\,\mathrm{V}) - (2\,\mathrm{V}) = 4\,\mathrm{V}$$

Again, you can verify that the voltage drops across the two resistors will sum to the total voltage drop across the terminals of the ideal battery.

Discussion: Modelling real batteries is not so different from modelling ideal batteries, since one only needs to include an internal resistance into the circuit. The key difference with a real battery is that the voltage across its real terminals depends on what is connected to the battery. In the example above, the battery has a voltage of $6\,\mathrm{V}$ across its (real) terminals when nothing is connected, but the voltage drops to $4\,\mathrm{V}$ when a $2\,\Omega$ resistor is connected.

Checkpoint 5-2

Suppose you would like to measure the change in voltage across the terminals of a battery by connecting a voltmeter in parallel with a battery. You would like the voltmeter (which acts as a resistor) to interfere with the measurement as little as possible. Would you choose a voltmeter with a high resistance, or a voltmeter with a low resistance?

A) High resistance.
B) Low resistance.
C) It doesn't matter if the voltmeter has a high or low resistance.

5.2 Kirchhoff's rules

Kirchhoff's rules correspond to concepts that we have already covered, but allow us to easily model more complex circuits, for instance, those where there is more than one path for the current to take. Kirchhoff's rules refer to "junctions" and "loops". Junctions and loops depend only on the shape of the circuit, and not on the components in the circuit. Figure 5.11 shows a circuit with no components in order to illustrate what is meant by a junction and a loop.

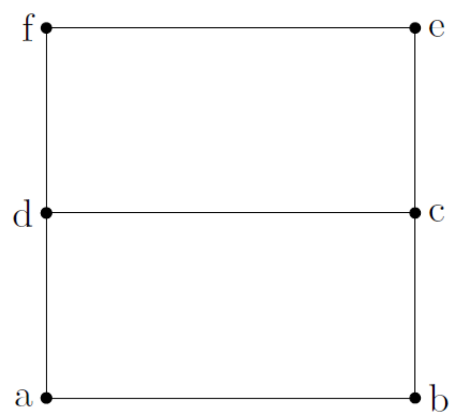

Figure 5.11: A circuit that has 3 loops and 2 junctions.

The locations at points d and c are considered "junctions", because there are more than 2 segments of wire connected to that point. The points at locations a, b, e and f only have two segments of wire connected to them. The circuit in Figure 5.11 thus has 2 junctions.

A loop is a closed path that one can trace around the circuit without passing over the same segment of wire twice. The circuit in Figure 5.11 has 3 such loops, which we can identify using the letters at the various nodes of the circuit:

1. $abcda$
2. $abcefda$
3. $dcefd$

Note that it does not matter where one starts on the loop, only that one can identify how many different loops are present in the circuit.

Checkpoint 5-3

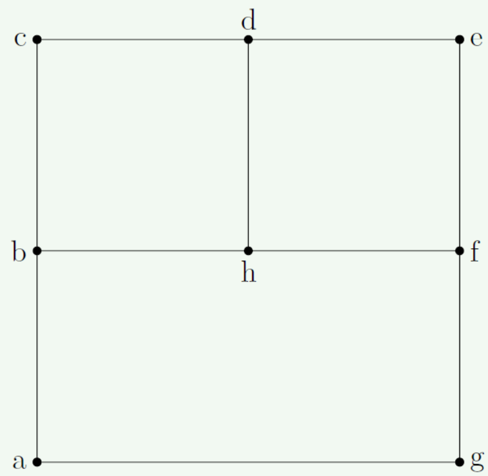

Figure 5.12

How many loops and junctions does the circuit in Figure 5.12 have?

A) The circuit has five loops and four junctions
B) The circuit has three loops and eight junctions
C) The circuit has seven loops and four junctions.
D) The circuit has four loops and four junctions.

5.2.1 Junction rule

The junction rule states that: **The current entering a junction must be equal to the current exiting a junction.**

This is in fact a simple statement about conservation of charge. If charges are flowing into a junction (from one or more segments of wire in that junction), then the same amount of charges must flow back out of the junction (through one or more different segments of wire).

Consider the junction illustrated in Figure 5.13, comprised of 5 segments of wire, and thus having 5 currents. As shown, currents I_1 and I_4 flow into the junction, whereas currents I_2, I_3 and I_5 all flow out of the junction.

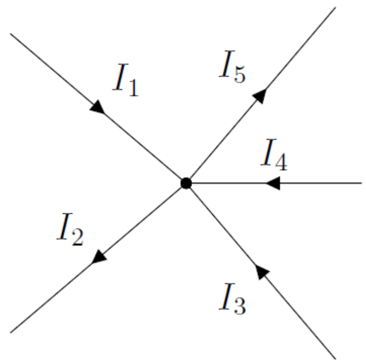

Figure 5.13: A junction with 5 segments and 5 currents.

The junction rule states that the current entering the junction must equal the current coming out of the junction. This allows us to relate the currents to each other in an equation:

$$\text{incoming currents} = \text{outgoing currents}$$
$$I_1 + I_4 = I_2 + I_3 + I_4$$

5.2.2 Loop rule

The loop rule states that: **The net voltage drop across a loop must be zero.**

This is a statement about conservation of energy, that we already noted in Example 5-1. Once you have identified a specific loop, if you trace a closed path around the loop, the electric potential must be the same at the end of the path as at the beginning of the path (since it is literally the same point in space). This means that if there is a voltage drop along the path (e.g. due to one or more resistors), then there must be equivalent voltage increases somewhere else on the path (e.g. due to one or more batteries). If this were not the case, it would be possible to have a path where charges could gain a net amount of energy by going around that path, which they could keep doing indefinitely and create an infinite amount of energy; instead, if charges gain potential energy in a battery, they must then loose exactly the same amount of energy inside one or more resistors along the path.

Figure 5.14 shows a loop (which could be part of a larger circuit) to which we can apply the loop rule. The loop contains two batteries, facing in opposite directions (which would not normally be a good use of batteries), as illustrated by the battery arrows.

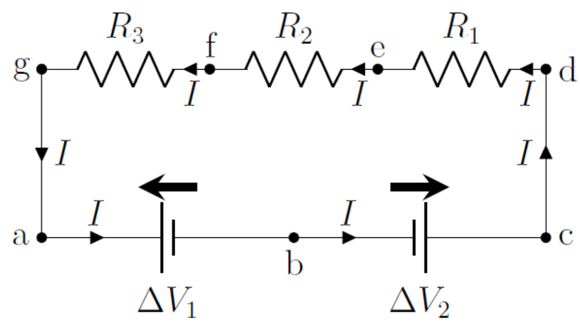

Figure 5.14: A loop with 2 batteries and 3 resistors.

The procedure for applying the loop rule is as follows:

1. Identify the loop, including starting position and direction.
2. Start at the beginning of the loop, and trace around the loop.
3. Each time a battery is encountered, **add the battery voltage if you are tracing the loop in the same direction as the corresponding battery arrow**, subtract the voltage otherwise.
4. Each time a resistor is encountered, **subtract the voltage across that resistor (RI, from Ohm's Law) if tracing the loop in the same direction as the current**, add the the voltage otherwise.
5. Once you have traced back to the starting point, the resulting sum must be zero.

To illustrate the procedure, we trace out the loop *abcedfga* in Figure 5.14. We thus start at point a and trace the loop in the counter-clockwise direction.

- Between points a and b we encounter a battery, and we are tracing in the **opposite direction of that battery's arrow**, so we subtract the voltage from that battery: $-\Delta V_1$.
- Between points b and c, we encounter a battery, and we are tracing in the **same direction as that battery's arrow**, so we add the voltage from that battery: $+\Delta V_2$.
- Nothing happens to the potential along the wire from c to d.
- Between points d and e, we encounter a resistor, and we are tracing in the **same direction as the current through that resistor**, so subtract the voltage across that resistor: $-R_1 I$).
- Similarly, we subtract the voltages across resistors R_2 and R_3, as we are tracing in the **same direction as the current through those resistors**: $-IR_2 - IR_3$.
- We are back at the beginning of the loop, so the terms must sum to zero.

We can now use the loop rule, which states that the sum of the above voltages must be zero:

$$-\Delta V_1 + \Delta V_2 - R_1 I - R_2 I - R_3 I = 0 \quad \text{(loop abcdefga)}$$

This equation then gives us a relation between the various quantities (current, resistors, battery voltages) in the circuit which can be used to model the circuit.

Checkpoint 5-4

1 Suppose the equation describing loop abcdefga was calculated from a different starting position and the loop was traced in the opposite direction. Would this prouce a different equation?

A) Yes, the equation would be incorrect if the loop is traced in the direction opposite to the flow of current.
B) Yes, the equation must start from the point a because the creator of the circuit assumes the person calculating current and voltage will begin at point a.
C) Yes, there is no incorrect starting point, but choosing to trace the circuit in the direction opposite to the current's flow would produce an incorrect equation.
D) No, there is no incorrect direction or starting point.

5.3 Applying Kirchhoff's rule to model circuits

In this section, we show how to model a circuit using Kirchhoff's rules. In general, one can consider a circuit to be fully modelled if one can determine the current in each segment of the circuit. We will show how one can apply the same procedure to model any circuit that contains batteries and resistors. The procedure is as follows:

1. Make a good diagram of the circuit.
2. Simplify any resistors that can easily be combined into effective resistors (in series or in parallel).
3. Make a new diagram with the effective resistors, showing battery arrows, and labelling all of the nodes so that loops can easily be described.
4. Make a **guess** for the directions of the current in each segment.
5. Write the junction rule equations.
6. Write the loop equations.
7. This will lead to N independent equations that one can solve for the N different currents in the circuit.
8. Once you have determined all of the currents, if some of them are negative numbers, switch the direction of those currents in the diagram (they will be negative if you guessed the direction incorrectly).

We will illustrate the procedure on the circuit shown in Figure 5.15, for which we would like to know the current through each resistor and each battery. The circuit contains 5 resistor (R_1-R_5), 2 real batteries (with ideal voltages ΔV_1 and ΔV_2), and 2 additional resistors to model the internal resistances of the batteries (r_1, r_2)

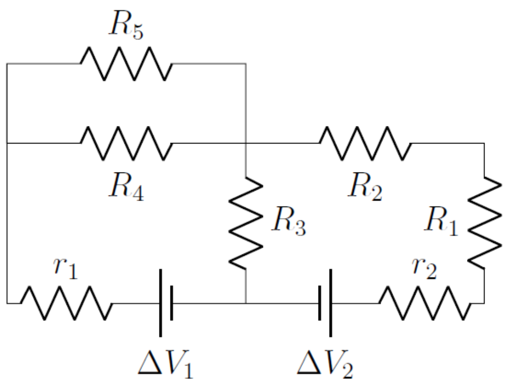

Figure 5.15: A circuit that can be simplified and then solved with Kirchhoff's rules.

Checkpoint 5-5

How many different currents are in the circuit shown in Figure 5.15 ?

A) 3
B) 4
C) 5
D) 6

Simplifying the resistors (step 2): In this circuit, resistors r_2, R_1 and R_2 are in series, so that they can be combined into an effective resistor, R_6:

$$R_6 = r_2 + R_1 + R_2$$

With this simplification, we obtain the circuit illustrated in Figure 5.16

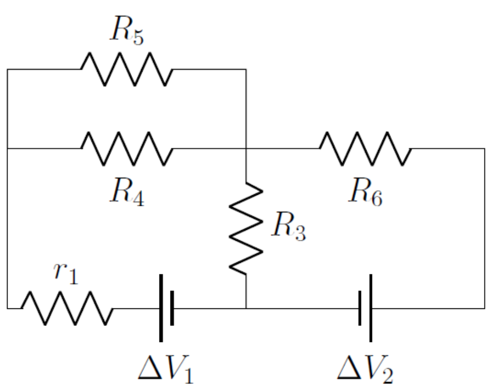

Figure 5.16: The resistors r_2, R_1 and R_2 in series from the circuit in Figure 5.15 have been combined into the effective resistor, R_6, to simplify the circuit.

Next, we note that resistors R_4 and R_5 are in parallel and can be easily combined into a resistor, R_7:

$$R_7 = \frac{R_4 R_5}{R_4 + R_5}$$

which leads to the circuit illustrated in Figure 5.17.

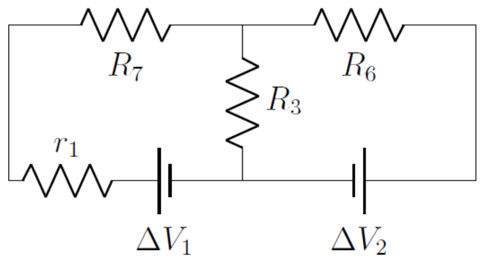

Figure 5.17: The resistors R_4 and R_5 in parallel from the circuit in Figure 5.16 have been combined into the effective resistor, R_7, to simplify the circuit.

Finally, we note that r_1 and R_7 are in series and can be combined into an effective resistor, R_8:

$$R_8 = r_1 + R_7 = r_1 + \frac{R_4R_5}{R_4 + R_5}$$

leading to the simplified circuit illustrated in Figure 5.18, which we have labelled with nodes and battery labels.

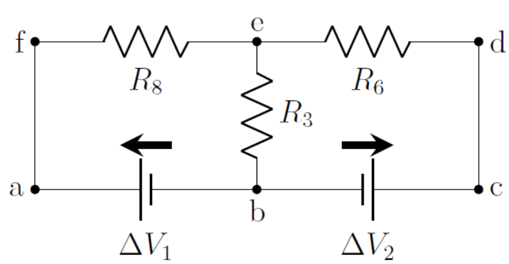

Figure 5.18: The resistors r_1 and R_7 in series from the circuit in Figure 5.17 have been combined into the effective resistor, R_8, to simplify the circuit.

Guessing the directions of the currents (step 4): Before we can write the equations from Kirchhoff's rules, we need to label the currents in the circuit diagram. In general, it is not always obvious in which way the currents will go, so we make a guess that we can fix later if we guessed wrong.

In order to guess the current directions, choose one point on the circuit and move along a segment. Label the current in that segment and continue moving through the circuit, splitting up the current when a junction is encountered. Make sure to only have one current per segment. We guess the currents as follows, referring to Figure 5.19:

- We start at point a and move upwards to point f. We will call the current in that segment, I_1.
- Since there is no junction, the current I_1 continues through the resistor R_8 to point e.
- There is a junction at point e, so we split the current I_1 into currents I_2 (towards point d), and I_3 (downwards to point b).
- We follow current I_2 first; I_2 flows from e to d, then down to c, through the battery ΔV_2, and to point b, where there is again junction.

- We follow current I_3, which just flows down to the junction at point b, where it "meets up" with current I_2.
- Currents I_2 and I_3 both flow into the junction at point b, and the current flowing out of the junction, through the battery ΔV_1, and towards point a is, again, I_1, since this current then flows up to point f.
- All segments of wire have a labelled current, so we are done guessing currents.

You can proceed in an analogous way for any circuit. The final circuit, with currents labelled, is shown in Figure 5.19:

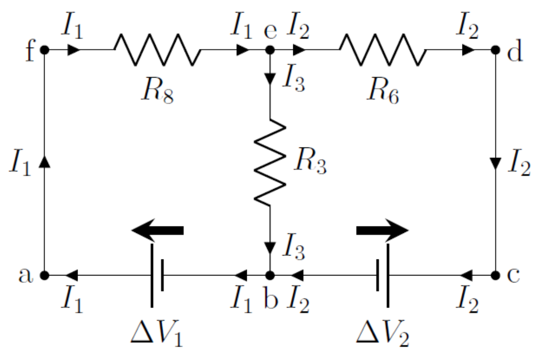

Figure 5.19: Final and labelled circuit diagram that is simplified from the one in Figure 5.15.

We can now proceed with using Kirchhoff's rules to solve for the values of the currents in the circuit. It is useful to note that there are 3 unknown currents in this circuit; we thus hope that Kirchhoff's rules will give us 3 independent equations.

Applying the junction rule (step 5): In the circuit from Figure 5.19, there are two junctions (at points b and e), so we will get two equations from the junction rule. To apply the junction rule, the sum of the currents coming into the junction must be equal to the currents going out of the junction:

$$\begin{aligned}
\text{incoming currents} &= \text{outgoing currents} \\
I_2 + I_3 &= I_1 && \text{(junction } b\text{)} \\
I_1 &= I_2 + I_3 && \text{(junction } e\text{)}
\end{aligned}$$

Note that the two equations are not independent (in fact, they are the same). It is generally the case that if there N junctions, one will obtain less than N independent equations (usually, $N-1$ equations will be independent). In this case, the two junctions only gave us one equation.

Applying the loop rule (step 6): This circuit contains 3 different loops: $abcdefa$, $abefa$, and $bcdeb$, which will lead to 3 equations from the loop rule. We expect that these equations will not be independent, since this would lead to 4 equations and 3 unknowns when combined with the junction rule equation. Let us start with the loop $abcdefa$:

- From a to b, we trace through the battery in the **opposite direction from the battery arrow**: $-\Delta V_1$.

- From b to c, we trace through the battery in the **same direction as the battery arrow**: $+\Delta V_2$.
- From c through d and through to e we go through the resistor R_6 in the **opposite direction from the current**, I_2, in that resistor: $+I_2R_6$.
- From e to f, we go through the go through the resistor R_8 in the **opposite direction from the current**, I_1, in that resistor: $+I_1R_8$.
- And we are back at the starting point, so the sum of the above terms is equal to zero.

which gives the equation:

$$-\Delta V_1 + \Delta V_2 + I_2R_6 + I_1R_8 = 0 \quad \text{(loop abcdefa)}$$

Similarly, for the loop $abefa$, we obtain:

$$-\Delta V_1 + I_3R_3 + I_1R_8 = 0 \quad \text{(loop abefa)}$$

and for loop $bcdeb$:

$$\Delta V_2 + I_2R_6 - I_3R_3 = 0 \quad \text{(loop bcdeb)}$$

Although it appears that we have obtained 3 additional equations, only two of these are independent. For example, if you sum the second and third equations (loops $abefa$, and $bcdeb$), you simply obtain the first equation (loop $abcdefa$). In general, if there are N different loops, one will obtain less than N independent equations (usually $N-1$ independent equations, as we did here).

At this point, we have 3 independent equations that we can solve for the 3 unknown currents[1]:

$$\begin{aligned} I_1 &= I_2 + I_3 && \text{(junction } e\text{)} \\ -\Delta V_1 + \Delta V_2 + I_2R_6 + I_1R_8 &= 0 && \text{(loop abcdefa)} \\ -\Delta V_1 + I_3R_3 + I_1R_8 &= 0 && \text{(loop abefa)} \end{aligned}$$

It is only a matter of some simple math to solve for the 3 unknowns from these 3 equations (which we carry out in the example below).

Example 5-3

Referring to the circuit in Figure 5.20, what is the voltage across the real terminal of the battery with ideal voltage ΔV_1 (the voltage between points a and b)? What is the current through resistor R_5?

Solution

[1]The 3 unknowns do not necessarily have to be the currents, and could be any combination of the currents, battery voltage and resistors. As long as there at most 3 unknown quantities, this circuit can be solved.

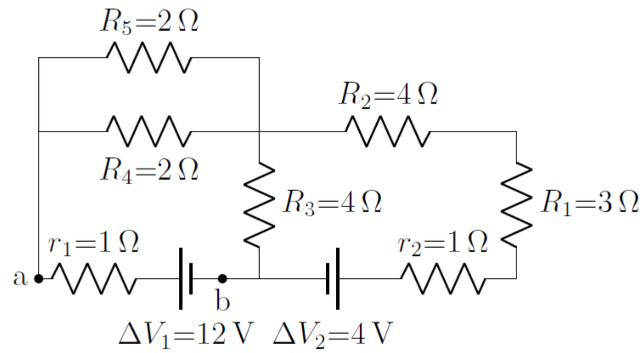

Figure 5.20: The same circuit as in Figure 5.15, with values filled in.

Since this circuit is the same that we just analyzed, we know that it can be simplified into the circuit shown in Figure 5.21, with resistors:

$$R_6 = r_2 + R_1 + R_2 = (1\,\Omega) + (3\,\Omega) + (4\,\Omega) = 8\,\Omega$$
$$R_8 = r_1 + \frac{R_4R_5}{R_4 + R_5} = (1\,\Omega) + \frac{(2\,\Omega)(2\,\Omega)}{(2\,\Omega) + (2\,\Omega)} = 2\,\Omega$$

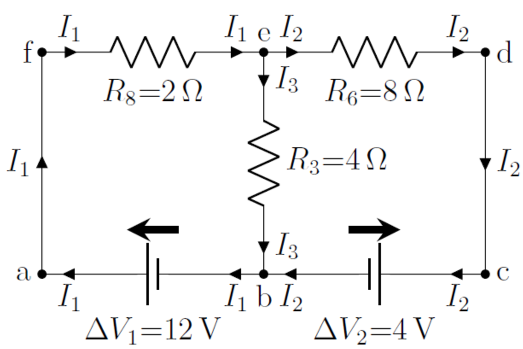

Figure 5.21: Simplified version of the circuit in Figure 5.20.

From above, we know that this leads to the following three equations:

$$\begin{aligned} I_1 &= I_2 + I_3 && \text{(junction } e\text{)} \\ -\Delta V_1 + \Delta V_2 + I_2R_6 + I_1R_8 &= 0 && \text{(loop abcdefa)} \\ -\Delta V_1 + I_3R_3 + I_1R_8 &= 0 && \text{(loop abefa)} \end{aligned}$$

In order to solve these types of equations, it is usually convenient to place the battery voltages on the right hand side, and the resistor voltages on the left hand side. Although it is generally bad practice to fill numbers into the equations before solving them, it is almost always a good idea when solving the N equations for the N currents. Furthermore, in order to make the equations legible, it is also useful to not write in the units (which is very bad practice in general!). Thus, filling in the values for the resistors and the battery voltages, moving the voltages to the right hand side, we obtain the following system of equations:

$$\begin{aligned}
I_1 - I_2 - I_3 &= 0 && \text{(junction } e\text{)}\\
2I_1 + 8I_2 &= 8 && \text{(loop abcdefa)}\\
2I_1 + 4I_3 &= 12 && \text{(loop abefa)}
\end{aligned}$$

Subtracting the second equation from the third equation (to eliminate I_1):

$$\begin{aligned}
4I_3 - 8I_2 &= 4\\
\therefore I_3 &= 1 + 2I_2
\end{aligned}$$

Substituting this into the junction equation:

$$\begin{aligned}
I_1 - I_2 - I_3 &= 0\\
I_1 - I_2 - 1 - 2I_2 &= 0\\
\therefore I_2 &= \frac{1}{3}(I_1 - 1)
\end{aligned}$$

Finally, substituting this into the equation from loop $abcdefa$, allows us to determine I_1 and the other two currents:

$$\begin{aligned}
2I_1 + 8I_2 &= 8\\
2I_1 + 8\left(\frac{1}{3}(I_1 - 1)\right) &= 8\\
\therefore I_1 &= \frac{16}{7} = 2.29\,\text{A}\\
\therefore I_2 &= \frac{1}{3}(I_1 - 1) = 0.43\,\text{A}\\
\therefore I_3 &= 1 + 2I_2 = 1.86\,\text{A}
\end{aligned}$$

In this case, the currents are all positive, so the diagram in Figure 5.21 is correct and we do not need to reverse the direction of any of the currents.

We can now determine the potential difference across the real terminals of the battery ΔV_1. The current through the battery is $I_1 = 2.29\,\text{A}$, which cause a voltage drop, ΔV_{r1}, across its internal resistance, r_1 of:

$$\Delta V_{r1} = I_1 r_1 = (2.29\,\text{A})(1\,\Omega) = 2.29\,\text{V}$$

The voltage across the real terminals of the battery is then:

$$\Delta V_{real} = \Delta V_1 - \Delta V_{r1} = (12\,\text{V}) - (2.29\,\text{V}) = 9.7\,\text{V}$$

The current through the resistor R_5 (Figure 5.20) requires a little more thought, since we calculated the current, I_1 through the effective resistor R_8, which we must now "break apart". Figure 5.22 shows the components of R_8.

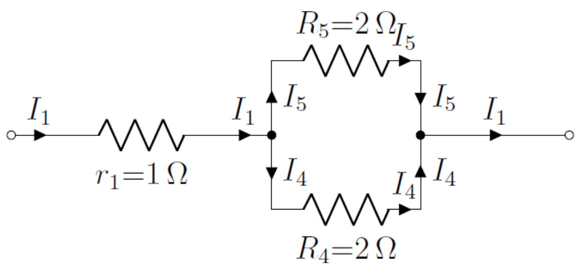

Figure 5.22: The components of the effective R_8 resistor from Figure 5.21. The current, I_1, coming from the battery goes through r_1 and then splits up.

The current I_1, that goes through the ΔV_1 battery also goes through the r_1 internal resistance of the battery. That current then splits up into currents, I_4 and I_5, to go through the resistors R_4 and R_5. Although it should be obvious that half of I_1 will go through each resistor (since these are equal), we can determine this from applying Kirchhoff's rules to the combination of resistors in Figure 5.22:

$$\begin{aligned} I_1 &= I_4 + I_5 && \text{(junction)} \\ I_5R_5 - I_4R_4 &= 0 && \text{(clockwise loop)} \end{aligned}$$

From the loop equation, we have:

$$I_5 = \frac{R_4}{R_5} I_4 = I_4$$

since $R_4 = R_5 = 2\,\Omega$. Since $I_4 = I_5$, the junction equation gives:

$$I_5 = \frac{1}{2} I_1 = 1.15\,\text{A}$$

By solving for I_4 and I_5, we have now determined all of the currents through all of the segments of the original circuit in Figure 5.20.

Discussion: In this example, we showed how one can use a simplified circuit to solve the current through the effective resistors in the simplified circuit. Once those currents are known, we showed that it is straightforward to determine the currents through individual resistors that have been combined into effective resistors.

Josh's Thoughts

Solving a circuit can be daunting, especially if the diagram is drawn in an unfamiliar way. While the circuits in this chapter are designed to be as easy to read as possible, many circuits are much more strange. For example, here is a circuit which you may come across:

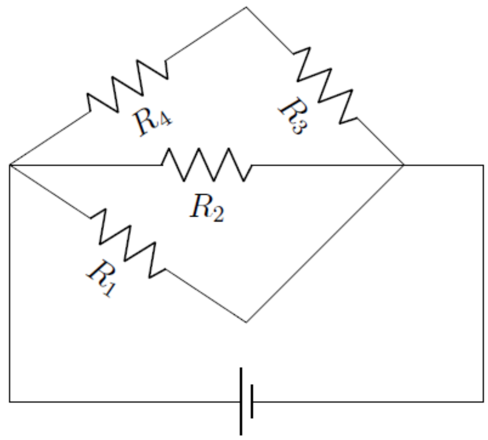

Figure 5.23: A weird looking circuit.

The circuit in Figure 5.23 May look like it is a difficult circuit to solve, but the diagram can be re-drawn to reveal the simplicity of the circuit:

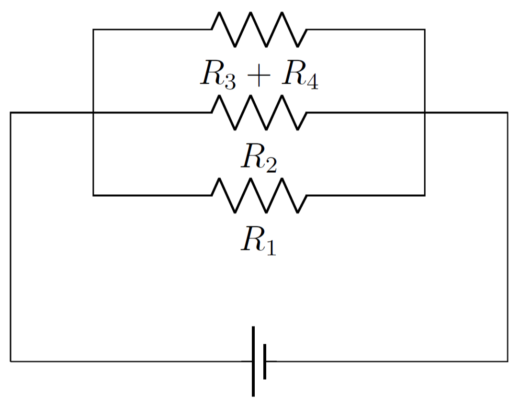

Figure 5.24: A much less weird looking circuit.

What used to be a strange kite shape is now just a parallel circuit, which can be further simplified by calculating the effective resistance:

$$R_{eff} = (R_1^{-1} + R_2^{-1} + (R_3 + R_4)^{-1})^{-1}$$

Which gives a series circuit with only one resistor:

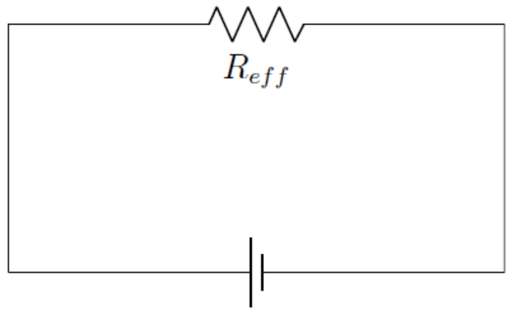

Figure 5.25: A simple circuit.

Circuits can be drawn in many unique or potentially confusing ways, but knowing how to read the circuit and re-draw it can help make the diagram more legible and the circuit easier to solve.

5.4 Measuring current and voltage

In this section, we describe how one can build devices to measure current and voltage. A device that measures current is called an "ammeter" and a device that measured voltage is called a "voltmeter". Nowadays, these are usually found within the same physical device (a "multimeter"), which can also measure resistance (by measuring voltage and current, resistance can easily determined). We will limit our description to the design of simple analogue ammeters and voltmeters.

As we will see in Chapter REF-MAGNETIC-FORCE-CHAPTER, it is straightforward to build a device that can measure very small amounts of current, by running the current through a coil in a magnetic field so that the coil can deflect a needle that indicates the amount of current. Such a device is called a "galvanometer" and is usually limited to measuring very small current (of order mA). In this section, we describe how one can use a galvanometer in order to build ammeters to measure large currents, and voltmeters.

5.4.1 The ammeter

An ammeter is built by placing a galvanometer in parallel with a "shunt" resistor, R_s. The shunt resistor is a small resistor that "shunts" (deflects) the current away from the galvanometer, so that most of the current goes through the shunt resistor. This is illustrated in Figure 5.26, which shows the galvanometer (circle with the G inside), the internal resistance of the galvanometer, R_G, and the shunt resistor, R_S. The actual ammeter would be contained in a box and have two connectors (shown as A and B in the figure).

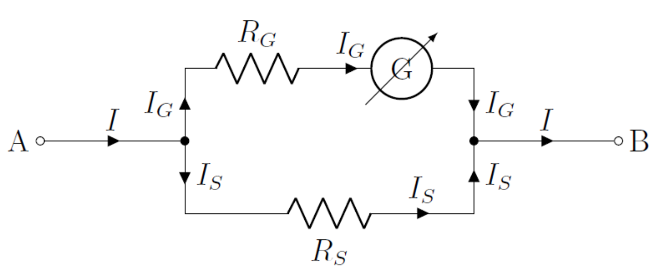

Figure 5.26: Constructing an ammeter from a galvanometer by placing a shunt resistor.

By modelling the ammeter, we can determine the total current, I, that we would like to measure using the values of the resistors and the and the current, I_G, measured by the galvanometer. Considering any of the two junctions, and a clockwise loop, we have:

$$\begin{aligned}
I &= I_G + I_S && \text{(junction)}\\
I_G R_G - I_S R_S &= 0 && \text{(clockwise loop)}\\
\therefore I_S &= \frac{R_G}{R_S} I_G\\
\therefore I &= I_G + _S = \left(1 + \frac{R_G}{R_S}\right) R_G
\end{aligned}$$

and indeed, we see that most of the current goes through the shunt (since R_S is chosen to be smaller than R_G). The ammeter, will have a total resistance, R_A, given by:

$$R_A = \frac{R_G R_S}{R_G + R_S}$$

In order to measure the current through a specific segment of a circuit, a ammeter must be placed in series with that segment (so that the current that we want to measure will pass through the ammeter). Figure 5.27 shows how to connect an ammeter (circle with the letter A) in order to measure the current through a resistor, R.

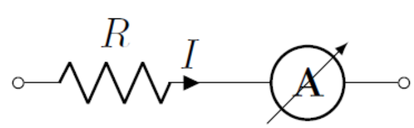

Figure 5.27: An ammeter is placed in series with a resistor to measure the current through the resistor.

5.4.2 The voltmeter

A voltmeter is constructed by placing a large resistor, R_V, in series with a galvanomenter (that has internal resistance R_G), as illustrated in Figure 5.28. The voltmeter is designed to measure the potential difference between the terminals of the voltmeter (labelled A and B in the Figure).

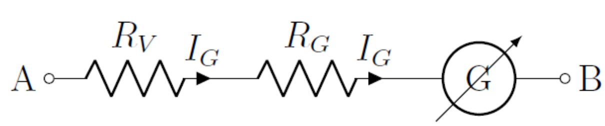

Figure 5.28: Constructing an voltmeter from a galvanometer by placing a resistor in series with the galvanometer.

Given the values of the resistors, and the current measured by the galvanometer, one can easily determine the potential difference between points A and B, since the current measured by the galvanometer goes directly through each resistor:

$$\Delta V = V_B - V_A = -I_G(R_V + R_G)$$

In order to measure a potential difference across a component, the voltmeter must be placed in parallel with the component. Figure 5.29 shows how to connect a voltmeter (circle with the letter V) in order to measure the voltage across a resistor, R.

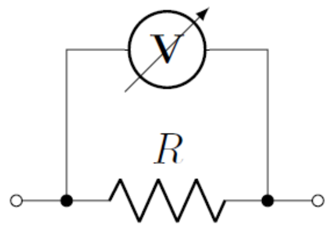

Figure 5.29: A voltmeter is placed in parallel with a resistor to measure the voltage across the resistor.

When using an ammeter or a voltmeter, you will notice that these usually have buttons or dials to choose the range of currents or voltages to be measured. All the dial does is change the value of the shunt or series resistor in order to maintain a given maximum current through the galvanometer. An ohmmeter, to measure resistance, is simply an ammeter with a built-in fixed potential difference (so that by measuring current across a known potential difference, the resistance of the component can be determined).

Example 5-4

Two resistors with a resistance of $1\,\mathrm{k\Omega}$ are placed in series with a $12\,\mathrm{V}$ battery. A voltmeter with a total resistance of $R_V = 10\,\mathrm{k\Omega}$ is used to measure the voltage across one of the resistors. What reading does the voltmeter show?

Solution

Since the two resistors have the same resistance, and are in series with the battery, when no voltmeter is connected, the voltage across either resistor is easily shown to be $6\,\mathrm{V}$. However, by connecting the voltmeter across one of the resistors, we modify the circuit, and we should expect the voltage that is read to be different than $6\,\mathrm{V}$ (can you tell if it will be larger or smaller?). The circuit, with the voltmeter connected is shown in Figure 5.30.

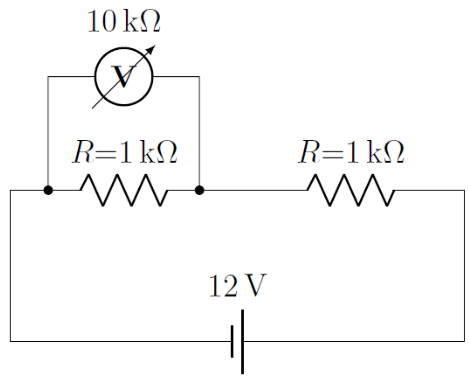

Figure 5.30: When using a voltmeter, the circuit is modified.

We can model this circuit quite easily by combining the voltmeter (modelled as a resistor) in parallel with one of the resistors:

$$R_{eff} = \frac{R_V R}{R_V + R} = \frac{(10\,\mathrm{k\Omega})(1\,\mathrm{k\Omega})}{(10\,\mathrm{k\Omega}) + (1\,\mathrm{k\Omega})} = \frac{10}{11}\mathrm{k\Omega} = 0.91\,\mathrm{k\Omega}$$

The sum of the voltage drops across the effective resistor and the other resistor must equal the potential difference across the battery (Kirchhoff's loop rule):

$$R_{eff}I + RI = \Delta V$$
$$\therefore I = \frac{\Delta V}{R_{eff} + R} = \frac{(12\,\mathrm{V})}{(0.91\,\mathrm{k\Omega}) + (1\,\mathrm{k\Omega})} = 6.29 \times 10^{-3}\,\mathrm{A}$$

The voltage drop across the effective resistor is the same as the reading on the voltmeter:

$$\Delta V_{voltmeter} = IR_{eff} = (6.29 \times 10^{-3}\,\mathrm{A})(0.91\,\mathrm{k\Omega}) = 5.7\,\mathrm{V}$$

and the voltmeter reads a smaller current than there would be without the voltmeter.

Discussion: In this example, we saw that by using a voltmeter to measure a voltage in a circuit, we actually disturb the circuit. By placing the voltmeter in parallel with one resistor, we created an effective resistor with a resistance that is lower than the resistance of either the voltmeter or the resistor. This lowered the total resistance of the circuit, which increased the current. A larger current through the second resistor (without the voltmeter) leads to a larger voltage drop than 6 V across that resistor. Thus, the voltage drop across the resistor with the voltmeter will be less than 6 V, as we found, since the two voltage drops need to add to 12 V.

In general, when using a voltmeter, one needs a voltmeter with a very high resistance in order to minimize the disturbance to the circuit (if the voltmeter has a high resistance, only a small amount of current will be shunted from the resistor). In practice, voltmeters have resistance that are typically of the order of 1 MΩ.

5.5 Modelling circuits with capacitors

Review Topics

- Section ?? on capacitors.

So far, we have modelled circuits where the current does not change with time. When a capacitor is included in a circuit, the current will change with time, as the capacitor charges or discharges. The circuit shown in Figure 5.32 shows an ideal battery[2] (ΔV), in series with a resistor (R), a capacitor (C, two vertical bars) and a switch (S) that is open.

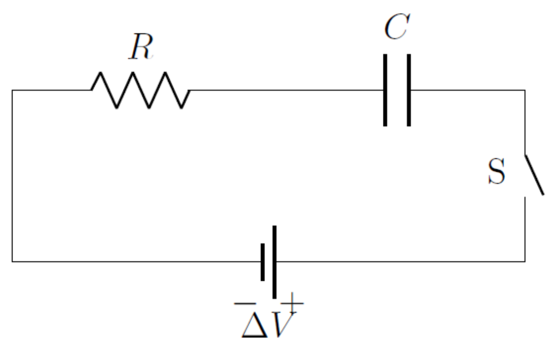

Figure 5.31: A simple circuit with a resistor, battery, and capacitor.

When the switch is open, no current flows through the circuit. If we assume that the capacitor has no charge on it, once we close the switch, current will start to flow and charges will accumulate on the capacitor. Electrons will leave the negative terminal of the battery, flow through the resistor and accumulate on the left side of the capacitor, which acquires a negative charge. This pushes electrons off of the right hand side of the capacitor, which then becomes positively charged. The electrons from the positive side of the capacitor then flow into the positive side of the battery, completing the circuit.

Eventually, the charges on the capacitor will build up to a point were they prevent the further flow of current. Once the left side of the capacitor is at the same potential as the left side of the battery, current will cease to flow. That is, eventually, the potential difference across the capacitor will be equal to that across the battery. The current is high when the switch is opened, but eventually goes down to zero as the capacitor charges, and the current is thus time-dependent.

We can model this simple circuit (with the switch closed) using Kirchhoff's loop rule. The sum of the voltages across each component must sum to zero:

$$\Delta V - IR - \frac{Q}{C} = 0$$

where we used the fact that the charge, Q, on a capacitor is related to the potential difference, ΔV_C, across the capacitor by $Q = C\Delta V_C$. The current, I, is the rate at which

[2] The model still holds for a real battery, since the internal resistance of the battery can just be included into the resistance of the resistor, R.

charges flow through the circuit, and is thus equal to rate at which charges accumulate on the capacitor:

$$I = \frac{dQ}{dt}$$

Substituting this into the loop equation, we obtain a separable differential equation for the charge on the capacitor as a function of time, $Q(t)$:

$$\begin{aligned}
\Delta V - IR - \frac{Q}{C} &= 0 \\
\Delta V - \frac{dQ}{dt}R - \frac{Q}{C} &= 0 \\
\Delta V - \frac{Q}{C} &= \frac{dQ}{dt}R \\
C\Delta V - Q &= RC\frac{dQ}{dt} \\
\therefore \frac{dt}{RC} &= \frac{dQ}{C\Delta V - Q}
\end{aligned}$$

This is similar to differential equations that we have solved previously (in fact, it is the same equation as in Example **??** where we looked at the effect of velocity-dependent drag). The solution to the equation, assuming that the switched is closed at $t = 0$, is given by an exponential:

$$Q(t) = C\Delta V\left(1 - e^{-\frac{t}{RC}}\right)$$

Thus, the charge on the capacitor starts at zero when the switch is closed, and grows asymptotically until it reaches a value of $Q = C\Delta V$, which corresponds to the capacitor having the same potential difference across it as the battery. The value $\tau = RC$ is called the "time constant" of the RC circuit, and corresponds to the time at which the capacitor will reach about $(1 - e^{-1}) = 63\%$ of its maximal charge. The current as a function of time is given by:

$$I(t) = \frac{dQ}{dt} = \frac{\Delta V}{R}e^{-\frac{t}{RC}}$$

and we can see that at time $t = 0$ the current is the same as if there were no capacitor present, and the current then decreases exponentially until it reaches zero.

5.6 Summary

Key Takeaways

Batteries are usually formed from a collection of electrochemical cells. Batteries provide a constant electric potential difference across their terminals, usually sustained by a chemical reaction, as long as the current through the battery is not too large (or the chemical reactions cannot be sustained). An ideal battery has no resistance and can be modelled as a simple potential difference in a circuit. A real battery includes an internal resistance and be modelled in a circuit as an ideal battery in series with a resistor. The voltage across the terminals of a real battery is equal to the voltage across the terminals of the ideal battery only when no current flows through the internal resistance.

Circuits are modelled using circuit diagram that include components (such as batteries and resistors) and wires. Wires are always modelled as having no resistance, since their resistance can be included by placing the appropriate resistor along the wire. The electric potential is always constant along a wire with no resistance. When modelling a circuit, **one always models the direction of conventional current**; that is, current is always indicated as the direction in which positive charges flow (even if in reality, it is negative electrons that flow in the opposite direction).

Circuits should be thought of in terms of conservation of energy. Components produce a potential difference between sections of wire. Batteries correspond to an increase in potential (if going from the negative to the positive terminal), whereas resistors corresponds to a decrease in potential (if going in the same direction as current through the resistor).

Kirchhooff's rules allow us to model complex circuits:

The junction rule states that: **The current entering a junction must be equal to the current exiting a junction.** This is a statement about conservation of charge. If charges are flowing into a junction, then the same amount of charges must flow back out of the junction per unit time.

The loop rule states that: **The net voltage drop across a loop must be zero.** This is a statement about conservation of energy indicating that as the potential energy of a positive charge increases as it goes through a battery, it will decrease by the same amount if it goes through a resistor that is connected in parallel to that battery.

In order to **apply the loop rule**, we strongly suggest using the following procedure, after having made a clear, labelled diagram showing battery arrows and currents in the circuit:

1. Identify the loop, including starting position and direction.
2. Start at the beginning of the loop, and trace around the loop.
3. Each time a battery is encountered, **add the battery voltage if you are tracing the loop in the same direction as the corresponding battery arrow**, subtract the voltage otherwise.

4. Each time a resistor is encountered, **subtract the voltage across that resistor (RI, from Ohm's Law) if tracing the loop in the same direction as the current**, add the the voltage otherwise.
5. Once you have traced back to the starting point, the resulting sum must be zero.

In general, we suggest the following procedure in order to use Kirchhooff's rules to model any circuit:

1. Make a good diagram of the circuit.
2. Simplify any resistors that can easily be combined into effective resistors (in series or in parallel).
3. Make a new diagram with the effective resistors, showing battery arrows, and labelling all of the nodes so that loops can easily be described.
4. Make a **guess** for the directions of the current in each segment.
5. Write the junction rule equations. In general, you will get $M-1$ independent equations for M loops.
6. Write the loop equations. In general, you will get $M-1$ independent equations for M loops.
7. This will lead to N independent equations that one can solve for the N different currents in the circuit.
8. Once you have determined all of the currents, if some of them are negative numbers, switch the direction of those currents in the diagram (they will be negative if you guessed the direction incorrectly).

Current and voltage measuring devices (ammeters and voltmeters, respectively) can be constructed from a galvanometer, which measures small currents. An ammeter is constructed by placing a small shunt resistor in parallel with the galvanometer so that most of the current passes through the shunt resistor. The resulting ammeter must be placed in series with a component in order to measure the current through that component.

A voltmeter is constructed by placing a resistor in series with the galvanometer in order to reduce the current through the galvanometer. The resulting voltmeter must be placed in parallel with a component in a circuit in order to measure the voltage across that component. Note that because voltmeters and ammeters have a non-zero resistance, they will affect the circuit once they are connected.

When a capacitor is placed in a circuit, the current in the circuit will no longer be constant in time. If a capacitor with capacitance, C, is placed in a series circuit with a battery and a resistor of resistance, R, the capacitor will charge until the voltage across the capacitor is equal to that across the battery. Once the capacitor is charged, current ceases to flow in the circuit. The charges on a capacitor accumulate with a rate that decays exponentially; it will take an infinite amount of time for the capacitor to become fully charged. It will be charged to about 63% of maximum charge after a period of time, $\tau = RC$, called the time constant of the capacitor.

Important Equations

Voltage:

$$V = IR$$

Junction Rule:

$$I_{in} = I_{out}$$

Loop Rule:

$$I_1 R_1 + I_2 R_2 + I_3 R_3 + I_n R_n + \Delta V_1 + \Delta V_2 + \Delta V_3 + \Delta V_n = 0$$

Resistance:

In a paralell circuit:

$$R_{total} = (R_1^{-1} + R_2^{-1} + R_3^{-1} + R_n^{-1})^{-1}$$

In a series circuit:

$$R_{total} = R_1 + R_2 + R_3 + R_n$$

5.7 Thinking about the material

Reflect and research

1. When did Galvani and Volta experiment with electric cells?
2. What is the largest voltage that Volta obtained with his voltaic pile?
3. How does one charge a rechargeable battery? What would the circuit look like?

To try at home

1. Research circuit diagrams of appliances you have at home.

To try in the lab

1. Propose an experiment to measure the change in current of an RC circuit as a capacitor builds up and releases charge.
2. Propose an experiment to measure the resistance of a voltmeter and compare your results with the manufacturer's.

5.8 Sample problems and solutions

5.8.1 problems

Problem 5-1: A simple RC circuit as shown in Figure **??** has a 9 V battery and a resistor with a resistance of 12 Ω. The switch is opened at $t = 0\,\mathrm{s}$. The current is then measured to be $I = 0.05\,\mathrm{A}$ at $t = 5\,\mathrm{s}$ after opening the switch. (Solution)

a) What is the capacitance of the capacitor?

b) What charge did the capacitor hold at $t = 2\,\mathrm{s}$?

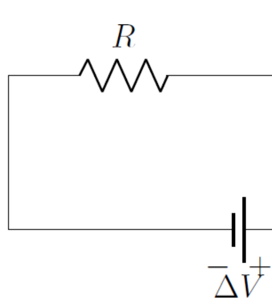

Figure 5.32: A simple circuit with a resistor, battery, and capacitor.

Problem 5-2: (Solution) A voltmeter with a resistance of $R_V = 20\,\mathrm{k\Omega}$ is attached to a circuit with a battery of unknown voltage and two resistors with a resistance of $R = 2.5\,\mathrm{k\Omega}$ as shown in 5.33. The voltmeter reads that the voltage drop over one of the resistors is $\Delta V_{vm} = 5.647\,\mathrm{V}$. What is the voltage drop, V_R, over each resistor when the voltmeter is removed from the circuit?

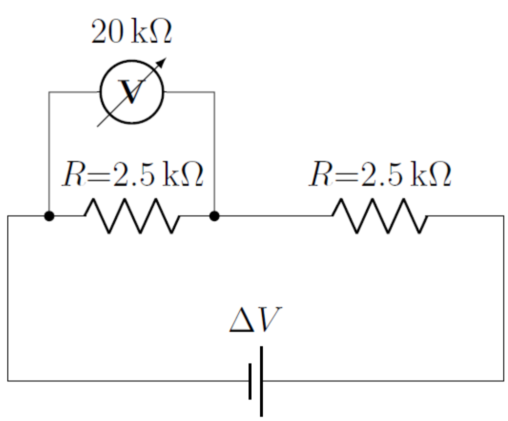

Figure 5.33: A circuit with a battery of unknown voltage.

5.8.2 Solutions

Solution to problem 5-2:

(a) In order to find the relationship between the capacitance and the current at some time t, we must use the following function:

$$I(t) = \frac{\Delta V}{R} e^{-\frac{t}{RC}}$$

This function describes the rate at which current decreases once the switch has been opened in the circuit. In this case, we have a measurement of current at $t = 5\,\mathrm{s}$, so we must solve for C in the time-dependent function:

$$\begin{aligned}
I(t) &= \frac{\Delta V}{R} e^{-\frac{t}{RC}} \\
I(t)\frac{R}{\Delta V} &= e^{-\frac{t}{RC}} \\
ln(I(t)\frac{R}{\Delta V}) &= -\frac{t}{RC} \\
ln(I(t)\frac{R}{\Delta V}) &= -\frac{t}{RC} \\
C &= -\frac{t}{Rln(I(t)\frac{R}{\Delta V})} \\
C &= -\frac{(5\,\mathrm{s})}{(12\,\Omega)ln((0.05\,\mathrm{A})\frac{12\,\Omega}{9\,\mathrm{V}})} \\
C &= 0.1539\,\mathrm{F}
\end{aligned}$$

The capacitance of the capacitor is $0.1539\,\mathrm{F}$.

(b) To find the charge stored in the capacitor at $t = 2\,\mathrm{s}$, we must use the time dependent equation for charge stored in a capacitor and our calculated capacitance:

$$\begin{aligned}
&Q(t) = C\Delta V(1 - e^{-\frac{-t}{RC}}) \\
&Q(2) = (0.1539\,\mathrm{F})(9\,\mathrm{V})(1 - e^{-\frac{-2\,\mathrm{s}}{(12\,\Omega)(0.1539\,\mathrm{F})}}) \\
&Q(2) = 0.9161\,\mathrm{C}
\end{aligned}$$

Which gives us the correct charge of $0.9161\,\mathrm{C}$ stored in the capacitor.

Solution to problem 5-2: In order to find the voltage in the battery, we must first find the current in the loop. We know the voltage drop over the voltmeter and resistor which are connected in paralell, so we must find their effective resistance, then use it to solve for current:

$$R_{eff} = \frac{1}{R_V^{-1} + R^{-1}}$$
$$R_{eff} = \frac{1}{(20\,\mathrm{k\Omega})^{-1} + (2.5\,\mathrm{k\Omega})^{-1}}$$
$$R_{eff} = 2.22\,\mathrm{k\Omega}$$

Now that we have the effective resistance as well as the voltage drop, we can solve for current:

$$I = \frac{\Delta V_{vm}}{R_{eff}}$$
$$I = \frac{5.647\,\mathrm{V}}{2.22\,\mathrm{k\Omega}}$$
$$I = 2.541\,\mathrm{mA}$$

Now that we have the current, we must sum the voltage drops in both resistors, which will give the battery's voltage:

$$\Delta V_{battery} = I(R_{eff} + R)$$
$$\Delta V_{battery} = (2.541\,\mathrm{mA})(2.222\,\mathrm{k\Omega} + 2.5\,\mathrm{k\Omega})$$
$$\Delta V_{battery} = 12\,\mathrm{V}$$

Now that we have the battery's voltage, we can solve for the voltage drop when the voltmeter is removed from the circuit. We know that the supplied voltage will be equal for each resistor, so the voltage drop over each resistor will be 6 V when the voltmeter is not present.