Physical Geology

Karla Panchuk

1st USask Edition
version 1, January 2019

UNIVERSITY OF SASKATCHEWAN

Physical Geology First University of Saskatchewan Edition (v1, January 2019). Adapted by Karla Panchuk from Physical Geology, 1st Edition by Steven Earle.

Chapters can also be downloaded from: http://openpress.usask.ca/physicalgeology/

Table of Contents

First USask Edition:

- Chapter 1. Introduction to Geology (1st U of S Ed.) Updated 2-12-2019
- Chapter 2. Origin of Earth & the Solar System (1st U of S Ed.) Updated 2-12-2019
- Chapter 3. Earth's Interior (1st U of S Ed.) Updated 2-12-2019
- Chapter 4. Plate Tectonics (1st U of S Ed.) Updated 10-1-2019
- Chapter 5. Minerals (1st U of S Ed.) Updated 10-1-2019
- Chapter 6. The Rock Cycle (1st U of S Ed.) Updated 10-1-2019
- Chapter 7. Igneous Rocks (1st U of S Ed.) Updated 10-1-2019
- Chapter 8: Weathering, Sediment, and Soil (1st U of S Ed.). Updated 10-1-2019
- Chapter 9. Sedimentary Rocks (1st U of S Ed.) Updated 10-1-2019
- Chapter 10. Metamorphism and Metamorphic Rocks (1st U of S Ed.) Updated 10-1-2019
- Chapter 11. Volcanism (1st U of S Ed.) Updated 10-1-2019
- Chapter 12. Earthquakes (1st U of S Ed.) Updated 10-1-2019
- Chapter 13. Geological Structures & Mountain Building (1st U of S Ed.) Updated 2-12-2019
- Chapter 14: Streams and Floods (1st U of S Ed.) Updated 17-1-2019
- Chapter 15: Mass Wasting (1st U of S Ed.) Updated 17-1-2019
- Chapter 16. Earth-System Change (1st U of S Ed.) Updated 10-1-2019
- Chapter 17. Glaciation (1st U of S Ed.) Updated 10-01-2019
- Chapter 18: Geological Resources (2nd Adapted Ed.). Updated 1-5-2017
- Chapter 19. Measuring Geological Time (1st U of S Ed.) Updated 10-1-2019

Chapters from Physical Geology (2015) Steven Earle:

- Chapter 14: Groundwater (1st Ed, Earle)
- Chapter 21. Geological History of Western Canada (1st Ed, Earle)

Cover image. View of Howe Sound a fjord on the south coast of B.C., Canada. This photo was taken from the South Summit of Stawamus Chief, a granodiorite pluton that is part of the coast mountains of B.C. Source: Joyce M. McBeth (2018) CC BY 4.0.

Preface to the First University of Saskatchewan Edition

Karla Panchuk

The First University of Saskatchewan Edition of *Physical Geology* is the product of several years' work iteratively adapting Steven Earle's original *Physical Geology* textbook. Edits since the spring of 2017 were supported financially through the University of Saskatchewan's Open Educational Resources Fund.

Key aspects of this latest version include:

- Expanded content on topics of particular importance to instructors at the University of Saskatchewan: Earth-system change, glaciation, and mass wasting
- Updated coverage of recent events and research;
- A consistent and transparent system for image attributions, making sources and permissions easy to trace.
- Additional images to support written content. The First University of Saskatchewan Edition contains approximately 580 figures, 40% of which were modified, added, or created as original works for this edition.

Thus far, this textbook (including previous adapted versions we've prepared) has been used by nearly a thousand students at the University of Saskatchewan, saving them tens of thousands of dollars in textbook costs. If you are considering adopting this version of this textbook in your courses or adapting it, please get in touch. We'd love to talk to you about what we've done so far and what we are planning for the next edition.

Karla Panchuk

January 2019

Preface to the Original Edition

This book was born out of a 2014 meeting of earth science educators representing most of the universities and colleges in British Columbia, and nurtured by a widely shared frustration that many students are not thriving in our courses because textbooks have become too expensive for them to buy. But the real inspiration comes from a fascination for the spectacular geology of western Canada and the many decades that I have spent exploring this region along with colleagues, students, family, and friends. My goal has been to provide an accessible and comprehensive guide to the important topics of geology, richly illustrated with examples from western Canada. Although this text is intended to complement a typical first-year course in physical geology, its contents could be applied to numerous other related courses.

As a teacher for many years, and as someone who is constantly striving to discover new things, I am well aware of that people learn in myriad ways, and that for most, simply reading the contents of a book is not one of the most effective ones. For that reason, this book includes numerous embedded exercises and activities that are designed to encourage readers to engage with the concepts presented, and to make meaning of the material under consideration. It is strongly recommended that you try the exercises as you progress through each chapter. You

should also find it useful, whether or not assigned by your instructor, to complete the questions at the end of each chapter.

Over many years of teaching Earth science I have received a lot of feedback from students. What gives me the most pleasure is to hear that someone, having completed my course, now sees Earth with new eyes, and has discovered both the thrill and the value of an enhanced understanding of how our planet works. I sincerely hope that this textbook will help you see Earth in a new way.

Steven Earle, Gabriola Island, 2015

Copyright

Unless otherwise noted, this book is released under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License also known as a CC BY-NC-SA 4.0 license. This means you are free to copy, redistribute, remix, transform, and build upon this book, as long as you share derivative materials using a CC BY-NC-SA 4.0 license. Under this license, anyone who redistributes this textbook can do so for free providing they properly attribute the book as follows: Physical Geology, First University of Saskatchewan Edition by Karla Panchuk used under a CC BY-NC-SA 4.0 international license. Additionally, if you redistribute this textbook, in whole or in part, in either a print or digital format, then you must retain on every physical and/or electronic page the following attribution: **https://openpress.usask.ca/physicalgeology/**

Acknowledgments

An open textbook for physical geology is something I had been considering ever since taking the *Introduction to Learning Technologies* course at the Gwenna Moss Centre for Teaching and Learning at the University of Saskatchewan. Adapting an open textbook is a far less daunting task than starting from scratch so I was excited to hear of the textbook *Physical Geology* by Steven Earle, written for the BCcampus Open Textbook project. Steven's original edition was a comprehensive and solid foundation on which to build this adapted work. Thanks to Amanda Coolidge of BCcampus for saving me an enormous amount of time by explaining how to modify the text and sending me the exported files from Steven's version of the textbook.

Many thanks go to Heather Ross and Nancy Turner at the Gwenna Moss Centre for their support and encouragement on this project and for discussions with them about open textbooks. The University of Saskatchewan Open Educational Resources Fund provided funding to support my work on this project. In-kind work and assistance on the project to match my time for this funding were provided by Joyce McBeth and Tim Prokopiuk of the Department of Geological Sciences.

This book has benefited from the work of numerous contributors at the University of Saskatchewan who have assisted with editing the document and providing new images to include in this edition. Tim Prokopiuk contributed edits and selected rock samples for me to photograph from the department's collection. Joyce McBeth provided numerous edits to this edition and adapted Chapters 14, 15, and 17. Lyndsay Hauber provided assistance with updates to image attributions for the chapter on plate tectonics. Donna Beneteau and Doug Milne of the College of Engineering, and Zoli Hajnal of Geological Sciences gave me a tour of

the Geological Engineering Rock Mechanics Facility, and helped me to photograph their experiments.

Image Sources

This project would not be possible without the generosity of many individuals and organizations who shared their work with a Creative Commons license or under other open licensing terms. The following is a list of valuable image resources, as much as it is an acknowledgement of contributions:

Roger Weller has made available thousands of his high-quality rock and mineral photographs through his website hosted by Cochise College, and granted permission for their non-commercial educational use. His photos have been used extensively throughout this project. Roger's usage stipulation has led to thoughtful discussions about what the appropriate way is to license derivative materials that make use of non Creative-Commons content. We have concluded that the best way to ensure that his wishes are respected is to license materials I make with his photographs as CC BY-NC-SA. This permits free sharing and remixing, but stipulates no commercial use, and that all derivative works must be shared with a non-commercial license.

James St. John is a geologist and paleontologist who has contributed (at the time of this writing) more than 59,000 high-quality geology-related photographs to the photo-sharing website Flickr. His photographs cover a wide range of rocks and minerals, and rarely has there been an image that I needed but couldn't find in his work. His Flickr account is remarkable for the abundance and quality of photographs, but also because he includes detailed descriptions of his images, making it possible for me to verify that an image is what I think it is, and gather useful background information. He has shared his images with a CC BY license, which I appreciate greatly because it allows me to combine them with content having more restrictive licenses.

The **U. S. Geological Survey** has contributed innumerable images to the public domain. The Hawaiian Volcano Observatory in particular is my go-to source for both the latest in volcano photos, and for fascinating historical images. Data and images from the USGS Earthquake Hazards Program Latest Earthquakes map have been invaluable.

I have used **NASA** images for views of Earth as much as I have for views of space and other planets. It is truly remarkable that in spite of the vast resources and expertise needed to acquire these photographs, they are free to view, use, and learn from.

Among the many teaching resources offered by **IRIS** (Incorporated Research Institutions for Seismology) are beautifully designed images for explaining earthquakes and seismology.

When all other sources failed, the odds were good that Robert Lavinsky (www.iRocks.com), Mike Norton, or Michael Rygel had contributed exactly the right photograph to Wikimedia Commons.

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-NC-SA 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 1. Introduction to Geology

Adapted by Karla Panchuk, University of Saskatchewan from *Physical Geology* by Steven Earle

Figure 1.1 Badlands in southern Saskatchewan. Erosion has exposed layers of rock going back more than 65 million years. *Source: Karla Panchuk (2017) CC BY-SA 4.0*

Learning Objectives

After reading this chapter, completing the exercises within it, and answering the *Review Questions* at the end, you should be able to:

- Explain what geology is, and why we study Earth.
- Describe the kinds of work that geologists do.
- Explain what is meant by *geological time*.
- Explain how the principle of uniformitarianism allows us to translate observations about Earth today into knowledge about how Earth worked in the past.
- Summarize the main idea behind the theory plate tectonics.

1.1 What Is Geology?

Geologists study Earth — its interior and its exterior surface, the rocks and other materials around us, and the processes that formed those materials. They study the changes that have occurred over the vast time-span of Earth's history, and changes that might take place in the near future.

Geology is a science, meaning that geological questions are investigated with deductive reasoning and scientific methodology. Geology is arguably the most interdisciplinary of all of the sciences because geologists must understand and apply other sciences, including physics, chemistry, biology, mathematics, astronomy, and more.

An aspect of geology that is unlike most of the other sciences is the role played by time — deep time — billions of years of it. When geologists study the evidence around them, they are often observing the results of are observing the results of events that took place thousands, millions, and even billions of years in the

past, and which may still be ongoing. Many geological processes happen at incredibly slow rates — millimetres per year to centimetres per year — but because of the amount of time available, tiny changes can result in expansive oceans forming, or entire mountain ranges being worn away.

Geology on a Grand Scale in the Canadian Rocky Mountains

The peak on the right of the photographs in Figure 11.2 is Rearguard Mountain, which is a few kilometres northeast of Mount Robson. Mount Robson is the tallest peak in the Canadian Rockies, at 3,954 m. The large glacier in the middle of the photo is the Robson Glacier. The river flowing from Robson Glacier drains into Berg Lake in the bottom right.

Many geological features are shown here. The rocks that these mountains are made of formed in ocean water over 500 million years ago. A few hundred million years later, the rocks were pushed east for tens to hundreds of kilometres, and thousands of meters upward in a great collision between Earth's tectonic plates.

Over the past two million years this area, like most of the rest of Canada, has been repeatedly covered by glaciers that scoured away rocks to form the valley to the left of Rearguard Mountain. The Robson Glacier itself is now only a fraction of its size during the Little Ice Age of the 15th to 18th centuries. And, like almost all other glaciers on Earth, it is now receding even more rapidly because of climate change. Figure 11.2 (right) taken around 1908 by the Canadian geologist and artist Arthur Philemon Coleman, gives an indication of how much the glacier has receded in the last hundred years.

Figure 11.2 Rearguard Mountain and Robson Glacier in Mount Robson Provincial Park, BC. Left: Robson Glacier today, retreating up the valley. Right: Robson Glacier circa 1908. *Sources: Left- Karla Panchuk (2017) CC BY-SA 4.0 with photo by Steven Earle (2015) CC BY 4.0. Right: A.P. Coleman (c. 1908) Public Domain.*

Geology is about understanding the evolution of Earth through time. It is about discovering resources such as metals and energy, and minimizing the environmental implications of our use of resources. It is about learning to mitigate the hazards of earthquakes, volcanic eruptions, and slope failures. All of these aspects of geology, and many more, are covered in this textbook.

1.2 Why Study Earth?

Why? Because Earth is our home — our *only* home for the foreseeable future — and in order to ensure that it continues to be a great place to live, we need to understand how it works. Another answer is that some of us can't help but study it because it's fascinating. But there is more to it than that.

- We can study rocks and the fossils they contain to understand the evolution of our environment and the life within it.
- We can learn to minimize our risks from earthquakes, volcanoes, slope failures, and damaging storms.
- We can learn how and why Earth's climate changed in the past, and use that knowledge to understand both natural and human-caused climate change.
- We rely on Earth for resources such as soil, water, metals, industrial minerals, and energy, and we need to know how to find these resources and exploit them sustainably.
- We can recognize how human activities have altered the environment, and learn how to prevent and sometimes repair the damage.
- We can use knowledge of Earth to understand other planets in our solar system, and those around distant stars.

The Importance of Geological Studies for Minimizing Risks to the Public

Figure 1.3 shows a slope failure that took place in January 2005 in the Riverside Drive area of North Vancouver. The steep bank beneath the house shown gave way, and a slurry of mud and sand flowed down. It destroyed another house below, and killed one person. The slope failure happened after a heavy rainfall, which is a common occurrence in southwestern B.C. in the winter.

Figure 1.3 Aftermath of a deadly debris flow in the Riverside Drive area of North Vancouver in January, 2005. *Source: The Province (2005), used with permission.*

A geological report written in 1980 warned the District of North Vancouver that the area was prone to slope failure, and that steps should be taken to minimize the risk to residents. Unfortunately, not enough was done in the intervening 25 years to prevent a tragedy.

1.3 What Do Geologists Do?

Geologists do a lot of different things. Many of the jobs are the things you would expect. Geologists work in the resource industry, including mineral exploration and mining, and exploring for and extracting sources of energy. They do hazard assessment and mitigation (e.g., assessment of risks from slope failures, earthquakes, and volcanic eruptions). They study the nature of the subsurface for construction projects such as highways, tunnels, and bridges. They use information about the subsurface for water supply planning, development, and management; and to decide how best to contain contaminants from waste.

Geologists also do the research that makes practical applications of geology possible. Some geologists spend their summers trekking through the wilderness to make maps of the rocks in a particular location, and collect clues about the geological processes that occurred there. Some geologists work in laboratories analyzing the chemical and physical properties of rocks to understand how the rocks will behave when forces act on them, or when water flows through them. Some geologists specialize in inventing ways to use complex instruments to make these measurements. Geologists study fossils to understand ancient animals and environments, and go to extreme environments to understand how life might have originated on Earth. Some geologists help NASA understand the data they receive from objects in space.

Geological work can be done indoors in offices and labs, but some people are attracted to geology because they like to be outdoors. Many geological opportunities involve fieldwork in places that are as amazing to see as they are interesting to study. Sometimes these are locations where few people have ever set foot, and where few ever will again.

Figure 1.4 Geologists at work on the island of Spitsbergen, part of the Svalbard archipelago. The islands are located in the Arctic Ocean north of Norway. *Source: Gus MacLeod (2007) CC BY-NC-ND 2.0*

1.4 We Study Earth Using the Scientific Method

There is no single method of inquiry that is specifically the scientific method. Furthermore, scientific inquiry is not necessarily different from serious research in other disciplines. The key features of serious inquiry are the following:

- Creation of a **hypothesis**. This is a tentative idea about how to explain a set of observations.
- Test the hypothesis. A hypothesis can be used to make predictions, and experiments can be run to see if those predictions are correct.

An Example of the Scientific Method at Work

Consider a field trip to the stream shown in Figure 1.5. Notice that the rocks in and along the stream are rounded off rather than having sharp edges. We might hypothesize that the rocks were rounded because as the stream carried them, they crashed into each other and pieces broke off.

Figure 1.5 Hypothesizing about the origin of round rocks in a stream. *Source: Steven Earle (2015) CC BY 4.0*

If the hypothesis is correct, then the further we go downstream, the rounder and smaller the rocks should be. Going upstream we should find that the rocks are more angular and larger. If we were patient we could also test the hypothesis by marking specific rocks and then checking back to see if those rocks have become smaller and more rounded as they moved downstream.

If the predictions turn out to be correct, we must still be careful about how much certainty to attach to our hypothesis. Although our hypothesis might seem to us to be the only reasonable explanation, someone could argue that we have the mechanism wrong, and the rocks weren't rounded by bumping into each other. If our experiment didn't specifically check for the mechanism (e.g., by looking to see if chips fall off the rocks and the rocks are made smoother) then we would have to acknowledge the possibility. We needn't abandon the hypothesis as a useful tool for making predictions, but it is necessary to be open to the possibility that other things might be going on. If someone demonstrates conclusively that our hypothesis is wrong, then we have to discard the hypothesis and come up with a better one.

A good hypothesis is testable. Someone might argue that an extraterrestrial organization creates rounded rocks and places them in streams when nobody is looking. There is no practical way to test this hypothesis to confirm it, and there is no way to prove it false. Even if we never see aliens at work, we still can't say they haven't been, because according to the hypothesis they only work when people aren't looking. Compare this to the original hypothesis that allows us to make testable predictions, such as rocks getting smaller and rounder downstream. Our original hypothesis gives us a way to see how realistic it is, whereas the alien hypothesis gives us no way to know if it makes sense or not.

Theories and Laws

Two other terms appear in discussions of the scientific method: **theory** and **law**. A **theory** starts out as a hypothesis, but over a long period of time and a great many tests, it has *never* come up short. That doesn't mean it never will, but the odds of that are very unlikely given our present (and conceivable future) state of knowledge. You may have heard someone dismiss an idea by saying it is "just a theory," but they are using the term incorrectly if they mean to say it's a wild and unproven guess.

A **law** is a description of a phenomenon rather than an explanation of it. For example, you could do thousands of tests by dropping an object with known mass and measuring its acceleration and the force with which it hits the ground. Again and again your results will yield the formula *force = mass x acceleration*. However, that doesn't mean you know what is responsible for the force accelerating it toward the ground. Yes, we say that gravity is pulling it toward the Earth's surface, but *why*? A law is true regardless of why a phenomenon happens as long as it describes the outcome of that phenomenon.

In geology there are three big ideas that are fundamental to the way we think about how Earth works. The ideas are like the sound track to a movie- sometimes we might not even notice them, but at the same time they affect our perception of what is happening. In the rest of this book these ideas may be mentioned explicitly in some cases, but in other cases it will be helpful for you to realize that they are relevant, even if they are not being discussed by name.

1.5 Three Big Ideas: Geological Time, Uniformitarianism, and Plate Tectonics

Geological Time (Deep Time)

Earth is approximately 4.57 billion years old (4,570,000,000 years), which is a long time for geological events to unfold and changes to happen. The changes themselves might be tiny. For example, over a year, a chemical reaction might eat away a few layers of atoms at the surface of a rock. But over time the changes accumulate and have a great impact. Over hundreds of millions of years the chemical reaction could cause a mountain range to crumble into grains of sand, and be swept away by rivers.

For geologists who study very, very slow processes, 10 million years might be a short time, and 1 million years might be trivial. For these geologists, intervals of 1 million years aren't even useful to consider, because the changes over that time are too small to see in the rocks that accumulated.

As you read through this book, keep in mind that the well of geologic time is indeed deep, and "ancient" is defined in a whole new way.

Expressing Geological Time in Numbers

Special notation is used for geological time because, as you might imagine, writing all those zeroes can become tiresome. Table 1.1 shows common abbreviations you will see throughout this book.

Table 1.1 Abbreviations Used to Describe Geological Time

Abbreviation	Meaning	Example
Ga	*giga annum* or billions of years	Earth is 4.57 Ga old.
Ma	*mega annum* or millions of years	Earth is 4,570 Ma old.
ka	*kilo annum* or thousands of years	The last glacial cycle ended 11,700 years ago, or 11.7 ka.

Expressing Geological Time Using the Geological Time Scale

The geological time scale (Figure 1.6) is a way of breaking down geological time according to important events in Earth's history. Time is divided into **eons**, **eras**, **periods**, and **epochs**, and these intervals are referred to by names rather than by years. Giving intervals of geologic time names rather than using numbers makes sense because we won't always know the age in years (the **absolute age**) of a rock or fossil, but we can place it in context based on our knowledge of the geological record. We can describe its **relative age** by saying that it is older than or younger than another rock or fossil**.**

Figure 1.6 Geologic Society of America Geologic Time Scale, 2012. *Source: Walker, J.D., Geissman, J.W., Bowring, S.A., and Babcock, L.E., compilers (2012) Geologic Time Scale v. 4.0: Geological Society of America, doi: 10.1130/2012.CTS004R3C. Note: See Additional Resources to download a large version of the timescale.*

The tricky thing about the geologic time scale is that the boundaries are always changing. As our knowledge of the absolute age of an event improves with new discoveries, it might be necessary to nudge a boundary earlier or later. Sometimes the original reason for defining a boundary no longer holds, but we agree to use it anyway. For example, the Phanerozoic Eon (the last 542 million years) is named for the time during which visible (*phaneros*) life (*zoi*) is present in the geological record, and its start was meant to mark the first appearance of these organisms. In fact, we now have evidence that large organisms — those that leave fossils visible to the naked eye — have existed longer than that, first appearing by 600 Ma at the latest.

An Early Definition of the Proterozoic

Notice that in Figure 1.6 the Proterozoic Eon precedes the Phanerozoic Eon. However, this was not always the case. Figure 1.7 shows an excerpt from a periodical published in 1879, in which the Proterozoic is defined as covering the Cambrian through Silurian. The author refers to "the most extreme adherents of the Murchisonian party in geology," a reference to the contentious assertion by Scottish geologist Roderick Murchison (1792-1871) that the Silurian Period should encompass the Cambrian and Ordovician periods as well.

THE ANNALS

AND

MAGAZINE OF NATURAL HISTORY,

INCLUDING

ZOOLOGY, BOTANY, AND GEOLOGY.

(BEING A CONTINUATION OF THE "ANNALS" CONDUCTED WITH LONDON AND CHARLESWORTH'S "MAGAZINE OF NATURAL HISTORY.")

CONDUCTED BY

ALBERT C. L. G. GÜNTHER, M.A., M.D., Ph.D., F.R.S.,
WILLIAM S. DALLAS, F.L.S.,
WILLIAM CARRUTHERS, F.R.S., F.L.S., F.G.S.,
AND
WILLIAM FRANCIS, Ph.D., F.L.S.

VOL. III.—FIFTH SERIES.

LONDON:
PRINTED AND PUBLISHED BY TAYLOR AND FRANCIS.
SOLD BY LONGMANS, GREEN, READER, AND DYER; SIMPKIN, MARSHALL, AND CO.; KENT AND CO.; WHITTAKER AND CO.; BAILLIÈRE, PARIS; MACLACHLAN AND STEWART, EDINBURGH; HODGES, FOSTER, AND CO., DUBLIN; AND ASHER, BERLIN.
1879.

Miscellaneous. 455

It will also be convenient to regard (with Professor Zittel) the first four families of the Rhabdophora as forming the artificial division of the *Monoprionida*, and the last four as constituting the similar division of the *Diprionida*.

Classification of the Lower Palæozoic Rocks.—The systems at present assigned to the Palæozoic age fall into two main groups—an older group, including the Cambrian, Ordovician, and Silurian systems, and a younger group, including the Devonian, Carboniferous, and Permian. The period during which the former were deposited may be denominated the *Lower Palæozoic or Proterozoic Age*; that in which the latter were laid down may be called the *Upper Palæozoic or Deuterozoic*. Broadly speaking, the Proterozoic rocks include all the sedimentary formations to which the name *Silurian* has at any time been applied by the most extreme adherents of the Murchisonian party in geology. With this extended interpretation the well-known generalization of Murchison that the Graptolites (Rhabdophora) are restricted to the Silurian system (Proterozoic period) still holds good. With one doubtful exception no Rhabdophora have hitherto been recorded from Deuterozoic rocks; their highest known limit

Figure 1.7 An excerpt from the periodical *The Annals and Magazine of Natural History* (1879) in which the name "Proterozoic" is assigned to the Cambrian, Ordovician, and Silurian periods instead of to the time preceding the Cambrian. *Source: Karla Panchuk (2017) CC BY 4.0. See Additional Resources to download the book.*

A Way To Think About Geological Time

A useful mechanism for understanding geological time is to scale it down into one year. The origin of the solar system and Earth at 4.57 Ga would be represented by January 1, and the present year would be represented by the last tiny fraction of a second on New Year's Eve. At this scale, each day of the year represents 12.5 million years; each hour represents about 500,000 years; each minute represents 8,694 years; and each second represents 145 years. Some significant events in Earth's history, as expressed on this time scale, are summarized in Table 1.2.

Table 1.2 Some Important Dates Expressed As If All of Geological Time Were Condensed Into One Year		
Event	**Approximate Date**	**Calendar Equivalent**
Formation of oceans and continents	4.5 - 4.4 Ga	first week of January
Evolution of the first primitive life forms	3.8 Ga	end of February
Formation of Saskatchewan's oldest rocks	3.4 Ga	end of March
Evolution of the first multi-celled animals	600 Ma	beginning of November
Animals first crawled onto land	360 Ma	end of November
Vancouver Island reached North America and the Rocky Mountains were formed	90 Ma	December 16
Extinction of the non-avian dinosaurs	65 Ma	December 18
Beginning of the Pleistocene ice age	2 Ma	10:10 p.m., December 31
Oldest radiocarbon date from people living in Canada (British Columbia)	13.8 ka	11:58 p.m., December 31
Earliest evidence of human activity in Saskatchewan	11.5 ka	48 seconds before midnight, December 31
The last of the glacial ice retreats from Saskatchewan	6 ka	41 seconds before midnight, December 31
Hudson's Bay Company establishes a permanent settlement at Cumberland House in northern Saskatchewan	243 years ago	2 seconds before midnight, December 31
Source: Karla Panchuk (2017) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0		

Uniformitarianism

Uniformitarianism is the notion that the geological processes occurring on Earth today are the same ones that occurred in the past. This is an important idea because it means that observations we make today about geological processes can be used to interpret and understand the rock record. While this idea might not seem remarkable today, it was ground breaking and even controversial for its time. Many people who heard about it for the first time thought about the age of the Earth in thousands of years, but uniformitarianism required them to think on timescales almost too vast to comprehend. For some, this implied questioning their most deeply held religious beliefs.

The Scottish geologist James Hutton initially presented the idea in 1785[1]. Charles Lyell, also a Scottish geologist, paraphrased this idea as "the present is the key to the past" in his book *Principles of Geology*[2]. This is how it is often described today.

To be clear, "the present is the key to the past" can be viewed as an oversimplification. Not all geological processes occurring today occurred at all times in the geological past. For example, some important chemical reactions that happened on Earth's surface today require abundant oxygen in the atmosphere, and

[1] See *Additional Resources* for a link to James Hutton's abstract. Note that the typeface prints an "s" like an "f."
[2] See *Additional Resources* for a link to *Principles of Geology.*

could not have occurred prior to Earth developing an oxygen-rich atmosphere. Conversely, there was a time in Earth's history when continents as we know them hadn't yet developed. Some events, such as devastating impacts by objects from space, have never been witnessed on the same scale by humans. We must be cognizant of the fact that conditions were different at different times in Earth's history, and take that into account when interpreting the rock record.

Despite the different past conditions on Earth as a whole, there still exist environments today where some of these conditions are present. These environments are like little samples of what Earth used to be like. This means we can still use present conditions to inform us about the past, but we have to think carefully about ways that such environments today differ from the ancient environments that no longer exist.

Plate Tectonics

It is only within the last 50 years or so that we have been able to answer questions like, "How did that mountain range get there?" and "Why do earthquakes happen where they do?" The theory of **plate tectonics**- the idea that Earth's surface is broken into large moving fragments, called **plates**- profoundly changed our perspective on how the Earth works. Figure 1.8 shows Earth's 15 largest tectonic plates, along with arrows indicating the plates' direction of motion, and how fast they go. (Longer arrows mean faster motion.) There are many more plates on Earth that are too small to show conveniently in Figure 1.8[3].

Figure 1.8 Earth's fifteen largest tectonic plates. Black arrows show the direction of plate motions. The length of the arrow indicates velocity. Red arrows show how plates move relative to each other. *Source: Steven Earle (2015) CC BY 4.0. Modified after U. S. Geological Survey (1996) Public Domain.*

[3] See *Additional Resources* for a link to a more detailed map of Earth's tectonic plates.

Prior to plate tectonics, we made observations but could only guess at mechanisms. It was like watching the hands on a clock and trying to guess what moves them. After plate tectonics it was like being able to open the clock and not only watch the gears turn, but realize for the first time that there are such things as gears. Plate tectonics not only explains why things have happened, but also allows us to predict what might happen in the future.

Plate tectonics is covered in more detail later, however the key point is that Earth's outer layer consists of rigid plates that are constantly interacting with each other as they move around the Earth. The boundaries of plates move away from each other in some places, collide in others, and sometimes just slide past each other (illustrated by the red arrows in Figure 1.8). The plates can move because they are floating on a layer of weak rock that deforms as the plates travel, much the same way the filling in a peanut butter and jelly sandwich allows you to slide the top layer of bread across the bottom layer.

Whether the plates move away from each other, collide, or just slide past each other determines things like the locations of mountain belts and volcanoes, where earthquakes happen, and the shapes and sizes of oceans and continents.

Chapter 1 Summary

The topics covered in this chapter can be summarized as follows:

1.1 What is Geology?

Geology is the study of Earth. It is an integrated science that involves the application of many of the other sciences. Geologists must take into account the fact that the geological features we see today may have formed thousands, millions, or even billions of years ago, and over very long time spans.

1.2 Why Study Earth?

Geologists study Earth out of curiosity and for other, more practical reasons, including understanding the evolution of life on Earth; searching for resources; understanding risks from geological events such as earthquakes, volcanoes, and slope failures; and documenting past environmental and climate changes so that we can understand how human activities are affecting Earth.

1.3 What Do Geologists Do?

Geologists work in the resource industry, and in efforts to protect the environment. Geologists work to minimize the risks from geological hazards (e.g., earthquakes), and to help the public understand those risks. Geologists investigate Earth materials in the field, in and in the lab.

1.4 We Study Earth Using the Scientific Method

Scientific inquiry requires a careful process of making a hypothesis and then testing it. If a hypothesis doesn't pass the test, it's time for a new one. A theory is a hypothesis that has been tested repeatedly and never failed a test. A law is a description of a natural process.

1.5 Three Big Ideas: Geological Time, Uniformitarianism, and Plate Tectonics

Geological time: Earth is approximately 4,570,000,000 years old; that is, 4.57 billion years or 4.57 Ga or 4,570 Ma. It’s such a huge amount of time that even extremely slow geological processes can have an enormous impact.

Uniformitarianism: Processes that occur today also occurred in the geologic past. We can use our observations of the present to understand the processes that shaped the Earth throughout its history.

Plate tectonics: Earth's surface is broken into plates that move and interact with each other. The interactions between these plates are key for understanding the mechanisms behind geologic processes.

Review Questions

1. How does the element of time make geology different from the other sciences, such as chemistry and physics?
2. List three ways in which geologists can contribute to society.
3. The following dates are written with the abbreviations Ga, Ma, and ka. Express the dates in years. (For example, 2.3 Ma = 2,300,000 years)

    a. 2.75 ka
    b. 0.93 Ga
    c. 4.2 Ma
    d. 0.2 ka.

4. Dinosaurs first appear in the geological record in rocks from about 215 Ma and then most became extinct at 65 Ma. What percentage of geological time does this represent?
5. If sediments typically accumulate at a rate of 1 mm/year, what thickness of sediment could accumulate over a period of 30 million years?
6. Does uniformitarianism mean that conditions on Earth are uniform, and never change?
7. Summarize the main idea behind plate tectonics.

Answers at the end of the chapter

References

Cottrell, M. (2006) *History of Saskatchewan*. Retrieved 26 August 2017 from http://esask.uregina.ca/entry/history_of_saskatchewan.html

Victoria University Library (2009) *A. P. Coleman Exhibition*. Retrieved 25 August 2017 from http://library.vicu.utoronto.ca/apcoleman/index.htm

Additional Resources

Download the *GSA Geologic Time Scale* in PDF form (487 KB): https://www.geosociety.org/documents/gsa/timescale/timescl.pdf

Download *The Annals and Magazine of Natural History* in PDF form (23 MB): http://openpress.usask.ca/physicalgeology/wp-content/uploads/sites/29/2017/08/Annals_Magazine_of_Natural_History-1.pdf

View the 7th edition of Charles Lyell's *Principles of Geology* from the Linda Hall Library Digital Collections at http://lhldigital.lindahall.org/cdm/ref/collection/earththeory/id/17524

View and dowload a detailed map of Earth's tectonic plates (3 MB): http://openpress.usask.ca/physicalgeology/wp-content/uploads/sites/29/2017/05/Plate_tectonics_map.gif

Answers to Review Questions

1. Geology requires that we consider vast amounts of time, and think about the effects that accumulate over thousands, millions, or even billions of years.
2. There are many ways that geologists contribute. Geologists provide information to reduce the risk of harm from hazards such as earthquakes, volcanoes, and slope failures; they play a critical role in the discovery of important resources; they contribute to our understanding of life and its evolution through paleontological studies; and they play a leading role in the investigation of climate change, past and present and its implications.
3. Ages in years
    a. 2.75 ka = 2,750 years
    b. 0.93 Ga = 930,000,000 years
    c. 14.2 Ma = 14,200,000 years
    d. 0.2 ka = 200 years.
4. 215 - 65 = 150 Ma. Since the age of the Earth is 4570 Ma, this represents 150/4,570 = 0.033 or 3.3% of geological time.
5. At 1 mm/y 30,000,000 mm of sediment would accumulate over that 30 million years. This is equivalent to 30,000 m or 30 km. Few sequences of sedimentary rock are even close to that thickness because most sediments accumulate at much lower rates, more like 0.1 mm/y. Also, over time the sediments are compressed.
6. No. Uniformitarianism means that we can use the processes we observe today to help us understand what happened in the past.
7. Plate tectonics is the idea that Earth's outer layer is broken into rigid plates. The plates move around and interact with each other along their margins.

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-NC-SA 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 2. Origin of Earth & the Solar System

by Karla Panchuk, University of Saskatchewan

Figure 2.1 Earthrise, October 12, 2015. The Lunar Reconnaissance Orbiter Camera captured images of the lunar surface with Earth in the background. *Source: NASA Lunar Reconnaissance Orbiter Science Team (2015) Public Domain.*

Learning Objectives

After reading this chapter and answering the review questions at the end, you should be able to:

- Explain the big bang theory for the origin of the universe.
- Explain how clouds of gas floating in space can turn into stars and planets.
- Describe the types of objects that are present in our solar system, and why they exist where they do.
- Explain how Earth got its layered structure, water, and atmosphere.
- Explain how the moon formed.
- Compare and contrast our solar system with other planetary systems.

The story of how Earth came to be is a fascinating contradiction. On one hand, many things had to go just right for Earth to turn out the way it did, and for life to develop. On the other hand, the formation of planets similar to Earth is an entirely predictable consequence of the physical and chemical processes taking place around stars. In fact, it has happened more than once.

This chapter starts Earth's story from the beginning — the *very* beginning — to explain why, for billions of years, generations of stars had to be born, then die explosive deaths before Earth could exist. How stars form and burn, and affect the objects around them are fundamental to Earth's story, as is the rough neighbourhood in which Earth spent its early years.

2.1 Starting with a Big Bang

According to the **big bang theory**, the universe blinked violently into existence 13.8 billion years ago. The big bang is often described as an explosion, but imagining it as an enormous fireball isn't accurate. The big bang started with a sudden expansion of energy and space from a single point. The kind of Hollywood explosion that might come to mind involves expansion of matter and energy *within* space, but during the big bang, energy, space, and matter were *created*. In Figure 2.2 the pointed base of the universe "vessel" represents the big bang. Time advances moving up the diagram. The vessel gets wider as time progresses, reflecting the expansion of the universe.

Figure 2.2 The big bang. The universe began 13.8 billion years ago as a rapid expansion of space, energy, and matter. It continues to expand. Left: Timeline of the universe. The point at the base of the "vessel" represents the moment of the big bang. The vessel gets wider as time progresses, representing the expansion of the universe. Right: Mollwiede projection of the cosmic microwave background, a "fog" from when the universe was still very dense. Temperature variations correspond to clumping of matter in the early universe. *Source: Karla Panchuk (2018) CC BY 4.0 modified after Ryan Kaldari (2006) Public Domain, derivative of NASA/WMAP Science Team (2006) Public Domain CMB map by NASA/WMAP Science Team (2006) Public Domain. See Appendix C for more attributions.*

You might wonder how a universe can be created out of nothing. Creating a universe out of nothing is mostly beyond the scope of this chapter, but there is a way to think about it. The particles that make up the universe have opposites that cancel each other out, similar to the way that we can add the numbers 1 and -1 to get zero (also known as "nothing"). As far as the math goes, having zero is exactly the same as having a 1 and a -1. It is also exactly the same as having a 2 and a -2, a 3 and a -3, two -1s and a 2, and so on. In other words, *nothing* is really the potential for *something* if you divide it into its opposite parts.

Composition of the Universe

In Figure 2.2, the "contents" of the vessel change as time progresses. A few minutes after the big bang, the universe was still too hot and dense to be anything but a sizzle of particles smaller than atoms. But as it expanded, it also cooled. Eventually particles that collided were able to stick together to form atoms, rather than being smashed apart again when other particles crashed into them. Those collisions produced hydrogen and helium, the most common elements in the universe.

For a long time after the big bang, clouds of hydrogen and helium atoms drifted about a dark universe. The "dark ages" (bottom of Figure 2.2) were a time when the ingredients for stars existed, but stars themselves did not. It took approximately 500 million years for enough hydrogen atoms to clump together in clouds to allow the first stars to form and begin to shine.

Looking Back to the Early Stages of the Big Bang

The notion of seeing the past is often used metaphorically when we talk about ancient events, but in this case it is meant literally. In our everyday experience, when we watch an event take place, we perceive that we are watching it as it unfolds in real time. In fact, this isn't true. To see the event, light from that event must travel to our eyes. Light travels very rapidly, but it does not travel instantly. If we were watching a digital clock 1 m away from us change from 11:59 a.m. to 12:00 p.m., we would actually see it turn to 12:00 p.m. three billionths of a second after it happened.

This isn't enough of a delay to cause us to be late for an appointment, but the universe is a very big place, and the "digital clock" in question is often much, much farther away. In fact, the universe is so big that it is convenient to describe distances in terms of **light years**, or the distance light travels in one year. What this means is that light from distant objects takes so long to get to us that we see those objects as they were at some considerable time in the past. For example, the star Proxima Centauri is 4.24 light years from the sun. If you viewed Proxima Centauri from Earth on January 1, 2018, you would actually see it as it appeared in early October 2013.

We now have tools that are powerful enough to look deep into space and see the arrival of light from early in the universe's history. Astronomers can detect light from approximately 380,000 years after the big bang is thought to have occurred. Physicists tell us that if the big bang happened, then particles within the universe would still be very close together at this time. They would be so close that light wouldn't be able to travel far without bumping into another particle and getting scattered in another direction. The effect would be to fill the sky with glowing fog, the "afterglow" from the formation of the universe.

In fact, this is exactly what we see when we look at light from 380,000 years after the big bang. The fog is referred to as the **cosmic microwave background** (or CMB), and it has been carefully mapped throughout the sky. In Figure 2.2, the colourful patch at the base of the diagram represents the fog that is measured today as the CMB. Figure 2.2 (right) is a CMB map of the universe in Mollweide projection. This is a projection that is used to show Earth's geography on a flat surface. In this case, the map of the CMB represents a sphere surrounding Earth rather than what's beneath our feet.

Colour variations in the CMB map represent temperature variations. These variations translate to differences in the density at which matter was distributed in the early universe. The red patches are the highest density regions and the blue patches are the lowest density. Higher density regions represent the eventual beginnings of stars and planets. The CMB map in Figure 2.2 has been likened to a baby picture of the universe.

The Universe is Still Expanding

The expansion that started with the big bang never stopped. It continues today, and we can see it happen by observing that large clusters of billions of stars, called **galaxies**, are moving away from us. (An exception is the Andromeda galaxy with which we are on a collision course.) The astronomer Edwin Hubble came to this conclusion when he observed that the light from other galaxies was red-shifted. The **red shift** is a consequence of the Doppler effect. This refers to how we see waves when the object that is creating the waves is moving toward us or away from us.

Before looking at the Doppler effect as it pertains to light, it can be useful to see how it works on something more tangible. The duckling swimming in Figure 2.3 is generating waves as it moves through

the water. It is generating waves that move forward as well as back, but notice that the ripples ahead of the duckling are closer to each other than the ripples behind the duckling. The distance from one ripple to the next is called the **wavelength**. The wavelength is shorter in the direction that the duckling is moving, and longer as the duckling moves away.

Figure 2.3 A duckling illustrates the Doppler effect in water. The ripples made in the direction the duckling is moving (blue lines) are closer together than the ripples behind the duckling (red lines). *Source: Karla Panchuk (2015) CC BY 4.0. Photo by M. Harkin (2013) CC BY 2.0*

When waves are in air as sound waves rather than in water as ripples, the different wavelengths manifest as sounds with different pitches — the short wavelengths have a higher pitch, and the long wavelengths have a lower pitch. This is why an observer will hear a change in the pitch of a car's engine as the car races past.

For light waves, wavelength translates to colour. In the spectrum of light that we can see, shorter wavelengths are on the blue end of the spectrum, and longer wavelengths are on the red end of the spectrum. In Figure 2.4, the longer or shorter wavelengths of the water ripples at the top of the diagram reflect the longer or shorter wavelengths of light in the visible spectra below. Does this mean that galaxies look red because they are moving away from us? No, but the colour we see is shifted toward the red end of the spectrum and longer wavelengths.

Notice that the sun's spectrum in the upper part of Figure 2.4 has black lines in it. The black lines are there because some colours are missing from the sun's light that reaches Earth. Different elements absorb light of specific wavelengths, and many of the black lines in Figure 2.4 represent colours that are absorbed by hydrogen and helium within the sun. This means the black lines are like a bar code that can tell us what a star is made of.

The lower spectrum in Figure 2.4 is the light coming from BAS11, an enormous cluster of approximately 10,000 galaxies located 1 billion light years away. The black lines represent the same elements as in the sun's spectrum, but they are shifted to the right toward the red end of the spectrum, because BAS11 is moving away from us as the universe continues to expand. To summarize, because almost all of the galaxies we can see have light that is red-shifted, it means they are all moving away from us. In fact, the farther away they are, the faster they are going. This is evidence that the universe is still expanding.

Figure 2.4 Red shift in light from the supercluster BAS11 compared to the sun's light. Black lines represent wavelengths absorbed by atoms (mostly hydrogen and helium). For BAS11 the black lines are shifted toward the red end of the spectrum compared to the sun. *Source: Karla Panchuk (2018) CC BY 4.0, spectra by Georg Wiora (2011) CC BY-SA 2.5*

2.2 Forming Planets from the Remnants of Exploded Stars

Only four elements account for 95% of Earth's mass: oxygen (O), magnesium (Mg), silicon (Si), and iron (Fe). Most of the remaining 5% comes from aluminum (Al), calcium (Ca), nickel (Ni), hydrogen (H), and sulphur (S). We know that the big bang made hydrogen, but where did the rest of the elements come from?

The answer is that the other elements were made by stars. Sometimes stars are said to "burn" their fuel, but burning is not what is going on within stars. The burning that happens when wood in a campfire is turned to ash and smoke is a chemical reaction — heat causes the atoms that were in the wood and in the surrounding atmosphere to exchange partners. Atoms group in different ways, but the atoms themselves do not change. What stars do is change the atoms. The heat and pressure within stars cause smaller atoms to smash together and fuse into new, larger atoms. For example, when hydrogen atoms smash together and fuse, helium is formed. Large amounts of energy are released when atoms fuse within stars, and this is what causes stars to shine. Stars can form large quantities of elements as heavy as iron during their normal burning process. Side reactions can form heavier elements in small amounts.

It takes larger stars to make elements as heavy as iron in large quantities. Our sun is an average star. After it uses up its hydrogen fuel to make helium, and some of that helium is fused to make small amounts of other elements, it will be at the end of its life. It will stop making new elements and will cool down and bloat until its middle reaches the orbit of Mars. In contrast, large stars end their lives in spectacular fashion. They explode as supernovae, casting off newly formed atoms into space, and triggering side reactions to make even more heavy atoms. It took many generations of stars creating heavier elements and casting them into space before heavier elements were abundant enough for planets like Earth to form.

Until recently, astronomers have only been able to see stars that already contain heavier elements in small amounts, but not the first-generation stars that started out before any of the heavier elements were produced. That changed in 2015 when it was announced that a distant galaxy called CR7 had been found that contained stars made only of hydrogen and helium. The galaxy is so far away that it shows us a view of the universe from approximately 800 million years after the big bang. Since then, more galaxies like CR7 have been discovered.

2.3 How to Build a Solar System

A **solar system** consists of a collection of objects orbiting one or more central stars. All solar systems start out the same way. They begin in a cloud of gas and dust called a **nebula**. Nebulae are some of the most beautiful objects that have been photographed in space. They have vibrant colours from the gases and dust they contain, and brilliant twinkling from the many stars that have formed within them (Figure 2.5). The

gas consists largely of hydrogen and helium, and the dust consists of tiny mineral grains, ice crystals, and organic particles.

Figure 2.5 The Pillars of Creation within the Eagle Nebula viewed in visible light (left) and near infrared light (right). Near infrared light captures heat from stars and allows us to view stars that would otherwise be hidden by dust. This is why the picture on the right appears to have more stars than the picture on the left. *Source: NASA, ESA, and the Hubble Heritage Team (STScI/AURA) (2015) Public Domain.*

Step 1: Collapse a Nebula

A solar system begins to form when a small patch within a nebula (small by the standards of the universe, that is) begins to collapse upon itself. Exactly how this starts isn't clear, although it might be triggered by the violent behaviour of nearby stars as they progress through their life cycles. Energy and matter released by these stars might compress the gas and dust in nearby neighbourhoods within the nebula.

Once it is triggered, the collapse of gas and dust within that patch continues for two reasons. One of those reasons is that gravitational force pulls gas molecules and dust particles together. But early in the process, those particles are very small, so the gravitational force between them isn't strong. So how do they come together? The answer is that dust first accumulates in loose clumps for the same reason dust bunnies form under the bed: static electricity. Given the role of dust bunnies in the early history of the solar system, one might speculate that an accumulation of dust bunnies poses a substantial risk to one's home (Figure 2.6). In practice, however, this is rarely the case.

Figure 2.6 Public service announcement. If you don't think housekeeping is important, then you don't understand the gravity of the situation. *Source: Karla Panchuk (2018) CC BY 4.0. Planets modified after NASA/JPL (2008) Public Domain.*

Step 2: Make a Disk with a Star at Its Centre

As the small patch within a nebula condenses, a star begins to form from material drawn into the centre of the patch, and the remaining dust and gas settle into a **protoplanetary disk** that rotates around the star. The disk is where planets will eventually form. Figure 2.7 (upper left) is an artist's impression of a protoplanetary disk, and Figure 2.7 (upper right) is an actual protoplanetary disk surrounding the star HL Tauri. Notice the dark rings within the HL Tauri protoplanetary disk. These are gaps formed by the collection of dust and debris by incipient planets, called **protoplanets**, as they orbit the star. There is an analogy for this in our own solar system, because the dark rings are akin to the gaps in the rings of Saturn (Figure 2.7, lower left), where moons can be found (Figure 2.7, lower right).

Figure 2.7 Protoplanetary disks and Saturn's rings. Upper left: Artist's impression of a protoplanetary disk containing gas and dust, surrounding a new star. Upper right: A photograph of the protoplanetary disk surrounding HL Tauri. The dark rings within the disk are thought to be gaps where newly forming planets are sweeping up dust and gas. Lower left: A photograph of Saturn showing similar gaps within its rings. The bright spot at the bottom is an aurora, similar to the northern lights on Earth. Lower right: a close-up view of a gap in Saturn's rings showing a moon as a white dot. *Source: Upper left- NASA/JPL-Caltech (2008) Public Domain; Upper right- ALMA (ESO/NAOJ/NRAO) (2014) CC BY 4.0; Lower left- NASA, ESA, J. Clarke (Boston University), and Z. Levay (STScI) (2005) Public Domain; Lower right- NASA/JPL/Space Science Institute (2005) Public Domain.*

Step 3: Build Some Planets

In general, planets can be classified into three categories based on what they are made of (Figure 2.8). **Terrestrial planets** are those planets like Earth, Mercury, Venus, and Mars that have a core of metal surrounded by rock. **Jovian planets** (also called **gas giants**) are those planets like Jupiter and Saturn that consist predominantly of hydrogen and helium. **Ice giants** are planets such as Uranus and Neptune that consist largely of water ice, methane (CH_4) ice, and ammonia (NH_3) ice, and have rocky cores. Often, the ice giant planets Uranus and Neptune are grouped with Jupiter and Saturn as gas giants; however, Uranus and Neptune are very different from Jupiter and Saturn.

Figure 2.8 Three types of planets. Jovian (or gas giant) planets such as Jupiter consist mostly of hydrogen and helium. They are the largest of the three types. Ice giant planets such as Uranus are the next largest. They contain water, ammonia, and methane ice, and have rocky cores. Terrestrial planets such as Earth are the smallest, and they have metal cores covered by rocky mantles. *Source: Karla Panchuk (2015) CC BY 4.0. See Appendix C for more attributions.*

These three types of planets are not mixed together randomly within our solar system. Instead they occur in a systematic way, with terrestrial planets closest to the sun, followed by the Jovian planets and then the ice giants (Figure 2.9). Smaller **solar system** objects follow this arrangement as well. The **asteroid belt** contains bodies of rock and metal. Bodies ranging from metres to hundreds of metres in diameter are classified as **asteroids**, and smaller bodies are referred to as **meteoroids**. In contrast, the **Kuiper belt** (*Kuiper* rhymes with *piper*), and the **Oort cloud** (*Oort* rhymes with *sort*), which are at the outer edge of the solar system, contain bodies composed of large amounts of ice in addition to rocky fragments and dust.

Part of the reason for this arrangement is the **frost line** (also referred to as the **snow line**). The frost line marks the division between the inner part of the protoplanetary disk closer to the sun, where it was too hot to permit anything but silicate minerals and metal to crystalize, and the outer part of the disk farther from the sun, where it was cool enough to allow ice to form. As a result, the objects that formed in the inner part of the protoplanetary disk consist largely of rock and metal, while the objects that formed in the outer part consist largely of gas and ice. The young sun blasted the solar system with raging **solar winds** (winds made up of energetic particles), which helped to drive lighter molecules toward the outer part of the protoplanetary disk.

Figure 2.9 Our solar system. Top: The solar system shown with distances to scale. Distances are in astronomical units (AU), where 1 AU is the average distance from Earth to the sun. The edge of the Kuiper belt extends to 50 AU (7.5 billion km), but this distance is minuscule compared to the size of the solar system as a whole, which extends to the edge of the Oort cloud, thought to be 15 trillion km away. Bottom: Solar system with the sun and planets to scale. The gas giants are the largest planets, followed by the ice giants, and then the terrestrial planets. Note that the planets in this diagram likely do not reflect the entire population of planets in our solar system because evidence suggests that large planets are present beyond the Kuiper belt. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0, Milky Way photo by R@pp (2017) CC BY-NC-SA 2.0, planet photographs courtesy of NASA. See Appendix C for planet photo sources and attributions.*

Rules of the Accretion Game

The objects in our solar system formed by **accretion**. Early in this process, particles collected in fluffy clumps because of static electricity. As the clumps grew larger, gravity became more important and collected clumps into solid masses, and solid masses into larger and larger bodies. If you were one of these bodies in the early solar system, and participating in the "accretion game" with the goal of becoming a planet, you would have to follow some key rules:

- Keep your velocity just right. If you move too fast and collide with another body, you both smash up and have to start again. If you move slowly enough, gravity will keep you from bouncing off each other and you can grow larger.
- Your distance from the Sun will determine how big you can get. If you are closer, there is less material for you to collect than if you are farther away.
- To begin with, you can only collect mineral and rock particles. You have to grow above a certain mass before your gravity is strong enough to hang onto gas molecules, because gas molecules are very light.
- As your mass increases, your gravity becomes stronger and you can grab material from farther away. The bigger you are, the faster you grow.

You would also have to watch out for some dangers:

- In the early stages of the game, the protoplanetary disk is turbulent, and you and other objects can get thrown into different orbits or at each other. This might be a good thing, or it might not, depending on how the rules above apply to you.

- If the game progresses to the point where there is no more material within your reach and you are not yet a planet, then it's game over.
- If you slow down too much (e.g., from bumping into other objects), you could spiral into the Sun (game over).
- If another planet gets big enough, it can:
    - Rip you apart and then swing the pieces around so fast that for the rest of the game you collide too hard with other pieces to grow any bigger (game over)
    - Fling you out of the solar system (game over)
    - Grab you for itself (game over)
    - Trap you in an orbit around it, turning you into a moon (game over, and incredibly humiliating)

Winners and Losers

The outcome of the game is evident in Figure 2.9. Today eight official winners are recognized, with Jupiter taking the grand prize, followed closely by Saturn. Both planets have trophy cases with more than 60 moons each, and each has a moon that is larger than Mercury. Prior to 2006, Pluto was also counted a winner, but in 2006 a controversial decision revoked Pluto's planet status. The reason was a newly formalized definition of a planet, which stated that an object can only be considered a planet if it is massive enough to have swept its orbit clean of other bodies. Pluto is situated within the icy clutter of the Kuiper belt, so it does not fit this definition.

Pluto's supporters have argued that Pluto should have been grandfathered in, given that the definition came after Pluto was declared a planet, but to no avail. Pluto has not given up, and on July 13, 2015, it launched an emotional plea with the help of the NASA's New Horizons probe. New Horizons sent back images of Pluto's heart (Figure 2.10). On closer inspection, Pluto's heart was discovered to be broken.

Figure 2.10 Photographs of Pluto. Left: The heart-shaped region called Tombaugh Regio is outlined. This region is named after Pluto's discoverer Clyde Tombaugh. Right: False-colour images show compositional variations in Tombaugh Regio. *Source: Karla Panchuk (2015) CC BY 4.0. Left photo- NASA/APL/SwRI (2015) Public Domain; Right photo- NASA/APL/SwRI (2015) Public Domain.*

The Accretion Game and the Solar System Today

The rules and dangers of the planet-forming game help to explain many features of our solar system today.

- Proximity to the sun explains why the terrestrial planets are so much smaller than the gas giant and ice giant planets.
- Mars is smaller than it should be, given the rule that distance from the sun determines how much material a body can accumulate, and this can be explained by its proximity to Jupiter. Jupiter's immense gravity interfered with Mars' ability to accrete. Further evidence of Jupiter's interference is the debris field that forms the asteroid belt. From time to time, Jupiter still flings objects from the asteroid belt out into other parts of the solar system, some of which have collided with Earth to catastrophic effect.
- The Kuiper belt is an icy version of the asteroid belt, consisting of fragments left over from the early solar system. The material in the Kuiper belt is scattered because of Neptune's gravity. From time to time, Jupiter interferes here as well, flinging Kuiper belt objects toward the sun and into

orbit. As these objects approach the sun, the sun causes dust and gas to be blasted from their surface, forming tails. We know these objects as comets.
- Comets may also come from the Oort cloud where gravitational forces from outside of the solar system can hurl objects from the Oort cloud toward the sun.

2.4 Earth's First 2 Billion Years

If you were to get into a time machine and visit Earth shortly after it formed (around 4.5 billion years ago), you would probably regret it. Large patches of Earth's surface would still be molten, which would make landing your time machine very dangerous indeed. If you happened to have one of the newer time-machine models with hovering capabilities and heat shields, you would still face the inconvenience of having nothing to breathe but a tenuous wisp of hydrogen and helium gas, and depending on how much volcanic activity was going on, volcanic gases such as water vapour and carbon dioxide. Some ammonia and methane might be thrown in just to make it interesting, but there would be no oxygen. Assuming you had the foresight to purchase the artificial atmosphere upgrade for your time machine, it would all be for naught if you materialized just in time to see an asteroid, or worse yet another planet, bearing down on your position. The moral of the story is that early Earth was a nasty place, and a time machine purchase is not something to take lightly.

Why was early Earth so nasty?

Earth Was Hot

Earth's heat comes from the decay of radioactive elements within Earth, as well as from processes associated with Earth's formation. Formation processes contributed heat in the following ways:

- Heat came from the thermal energy already contained within the objects that accreted to form Earth.
- Heat came from collisions. When objects hit Earth, some of the energy from their motion went into deforming Earth, and some of it was transformed into heat. Clap your hands vigorously to experience this on a much smaller (and safer!) scale.
- As Earth became larger, its gravitational force became stronger. This increased Earth's ability to draw objects to it, but it also caused the material making Earth to be compressed, rather like Earth giving itself a giant gravitational hug. Compression causes materials to heat up.

Heating had an important consequence for Earth's structure. As Earth grew, it collected a mixture of rocky silicate mineral grains as well as iron and nickel. These materials were scattered throughout Earth. That changed when Earth began to heat up: it got so hot that the metals melted and trickled down through the rocky silicate material toward Earth's centre, becoming Earth's core. The silicate material became Earth's crust and mantle. In other words, Earth unmixed itself. The separation of silicate minerals and metals into a rocky outer layer and a metallic core, respectively, is called **differentiation**. Friction from metal melts moving through Earth caused it to heat up even more.

Earth's high temperature early in its history also means that early tectonic processes were accelerated compared to today, and Earth's surface was more geologically active.

Earth Was Bombarded by Objects from Space

Although Earth had swept up a substantial amount of the material in its orbit as it was accreting, unrest within the solar system caused by changes in the orbits of Saturn and Jupiter was still sending many large objects on cataclysmic collision courses with Earth. The energy from these collisions repeatedly melted and even vaporized minerals in the crust, and blasted gases out of Earth's atmosphere. Very old scars from these collisions are still detectable, although we have to look carefully to see them. For example, the oldest

impact site discovered is the 3 billion year old Maniitsoq "crater" in west Greenland, although there is no crater to see. What is visible are rocks that were 20 km to 25 km below Earth's surface at the time of the impact, but which nevertheless display evidence of deformation that could only be produced by intense, sudden shock.

The evidence of the very worst collision that Earth experienced is not subtle at all. In fact, you have probably looked directly at it hundreds of times already, perhaps without realizing what it is. That collision was with a planet named Theia, which was approximately the size of Mars. Not long after Earth formed, Theia struck Earth (Figure 2.11). When Theia slammed into Earth, Theia's metal core merged with Earth's core, and debris from the outer silicate layers was cast into space, forming a ring of rubble around Earth. The material within the ring coalesced into a new body in orbit around Earth, giving us our moon. Remarkably, the debris may have coalesced in 10 years or fewer! This scenario for the formation of the moon is called the **giant impact hypothesis.**

Figure 2.11 Artist's impression of a collision between planets. A similar collision between Earth and the planet Theia might have given us our moon. *Source: NASA/ JPL-Caltech (2009) Public Domain.*

Today's Atmosphere Took a Long Time to Develop

Earth's first experiment with having an atmosphere did not succeed. It started out with a thin veil of hydrogen and helium gases that came with the material it accreted. But hydrogen and helium are very light gases, and they bled off into space.

Earth's second experiment with having an atmosphere went much better. Volcanic eruptions built up the atmosphere by releasing gases. The most common volcanic gases are water vapour and carbon dioxide (CO_2), but volcanoes release a wide variety of gases. Other important contributions include sulphur dioxide (SO_2), carbon monoxide (CO), hydrogen sulphide (H_2S), hydrogen gas, and methane (CH_4). Meteorites and comets also brought substantial amounts of water and nitrogen to Earth. It is not clear what the exact composition of the atmosphere was after Earth's second experiment, but carbon dioxide, water vapour, and nitrogen were likely the three most abundant components.

One thing we can say for sure about Earth's second experiment is that there was effectively no **free oxygen** (O_2, the form of oxygen that we breathe) in the atmosphere. We know this in part because prior to 2 billion years ago, there were no rocks stained red from oxidized iron minerals. Iron minerals were present, but not in oxidized form. At that time, O_2 was produced in the atmosphere when the sun's ultraviolet rays split water molecules apart. However, chemical reactions removed the oxygen as quickly as it was produced.

It wasn't until well into Earth's third experiment — life — that the atmosphere became oxygenated. Photosynthetic organisms used the abundant CO_2 in the atmosphere to manufacture their food, and released O_2 as a by-product. At first all of the oxygen was consumed by chemical reactions as before, but eventually

the organisms released so much O_2 that it overwhelmed the chemical reactions. Oxygen began to accumulate in the atmosphere, although present levels of 21% oxygen didn't occur until about 350 million years ago. Today the part of our atmosphere that isn't oxygen consists largely of nitrogen (78%).

The oxygen-rich atmosphere on our planet is life's signature. If geologic processes were the only ones controlling our atmosphere, it would consist mostly of carbon dioxide, like the atmosphere of Venus. It is an interesting notion (or a disconcerting one, depending on your point of view) that for the last 2 billion years the light reflected from our planet has been beaming a bar code out to the universe, similar to the ones in Figure 2.4, except ours says "oxygen." For 2 billion years, our planet has been sending out a signal that could cause an observer from another world to say, "That's odd… I wonder what's going on over there."

2.5 Are There Other Earths?

As of January 2018, 6,355 possible **exoplanets**- extrasolar planets, or planets outside of our solar system- have been detected by preliminary tests. Further tests have confirmed that 3,726 of those candidates are indeed planets. If "other Earths" are defined as planets where we could walk out of a spaceship with no equipment other than a picnic basket, and enjoy a pleasant afternoon on a grassy slope near a stream, then it remains to be seen whether any of these planets fit the description. On the other hand, if "other Earths" refers to rocky worlds approximately Earth's size, and orbiting within their star's **habitable zone** (the zone in which liquid water, and potentially life, can exist), then there is cautious optimism that we have found at least 53 such worlds.

Part of the uncertainty about the 53 possible Earth-like worlds is related to their composition. We don't yet know their composition; however, it is tempting to conclude that they are rocky because they are similar in size to Earth. Remember the rules of the accretion game: you can only begin to collect gas once you are a certain size, and how much matter you collect depends on how far away from the sun you are. Given how large our gas giant and ice giant planets are compared to Earth, and how far away they are from the sun, we would expect that a planet similar in size to Earth, and a similar distance from its star, should be rocky.

But it isn't quite as simple as that. We are finding that the rules to the accretion game can result in planetary systems very different from our own. For example, in the planetary systems we have observed, it is common to have planets larger than Earth orbiting closer to their star than Mercury does to the sun. Planets as large as Jupiter are rare, and where large planets do exist, they are much closer to their star than Jupiter is to the sun. To summarize, we need to be cautious about drawing conclusions from our own solar system, just in case we are basing those conclusions on something truly unusual.

On the other hand, the seemingly unique features of our solar system would make planetary systems like ours difficult to spot. One of the ways exoplanets are detected is by measuring the brightness of stars, and looking for tiny variations in brightness that could be caused by a planet passing between the star it orbits and the instrument observing the star. Small planets are harder to detect because they block less of a star's light than larger planets. Larger planets farther from a star, like our gas giant planets, are difficult to spot because they don't go past the star as frequently. For example, Jupiter goes around the sun once every 12 years. If someone were observing our solar system, they might have to watch for 12 years to see Jupiter go past the sun once. For Saturn, they might have to watch for 30 years.

If Habitable Zone Planets Are Terrestrial, Could We Live There?

The operational definition of "other Earths" involving a terrestrial composition, a size constraint of one to two times that of Earth, and location within a star's habitable zone, does not preclude worlds incapable of supporting life as we know it. By those criteria, Venus is an "other Earth," albeit right on the edge of the habitable zone for our sun. Venus is much too hot for us, with a constant surface temperature of 465°C (lead melts at 327°C). Its atmosphere is almost entirely carbon dioxide, and the atmospheric pressure at its surface is 92 times higher than on Earth. Any liquid water on its surface boiled off long ago. Yet the characteristics that make Venus a terrible picnic destination aren't entirely things we could predict from its

distance from the sun. They depend in part on the geochemical evolution of Venus- at one time Venus might have been a lot more like a youthful Earth. These are the kinds of things we won't know about until we can look carefully at the atmospheres and compositions of habitable-zone exoplanets.

Chapter 2 Summary

The topics covered in this chapter can be summarized as follows:

2.1 Starting With a Big Bang

The universe began 13.8 billion years ago when energy, matter, and space expanded from a single point. Evidence for the big bang is the cosmic "afterglow" from when the universe was still very dense, and red-shifted light from distant galaxies, which tell us the universe is still expanding.

2.2 Forming Planets from the Remnants of Exploded Stars

The big bang produced hydrogen, helium, but heavier elements come from nuclear fusion reactions in stars. Large stars make elements such as silicon, iron, and magnesium, which are important in forming terrestrial planets. Large stars explode as supernovae and scatter the elements into space.

2.3 How to Build a Solar System

Solar systems begin with the collapse of a cloud of gas and dust. Material drawn to the centre forms a star, and the remainder forms a disk around the star. Material within the disk clumps together to form planets. In our solar system, rocky planets are closer to the sun, and ice and gas giants are farther away. This is because temperatures near the sun were too high for ice to form, but silicate minerals and metals could solidify.

2.4 Earth's First 2 Billion Years

Early Earth was heated by radioactive decay, collisions with bodies from space, and gravitational compression. Heating caused molten metal to sink to Earth's centre and form a core, and silicate minerals to form the mantle and crust. A collision with a planet the size of Mars knocked debris into orbit around Earth, and the debris coalesced into the moon. Earth's atmosphere is the result of volcanic degassing, contributions by comets and meteorites, and photosynthesis.

2.5 Are There Other Earths?

The search for exoplanets has identified 53 planets that are similar in size to Earth and within the habitable zone of their stars. These are thought to be rocky worlds like Earth, but the compositions of these planets are not known for certain.

Review Questions

1. How can astronomers view events that happened in the universe's distant past?

2. In the image of three spectra at right, one is from the sun, and the other two are from galaxies. One of the galaxies is the Andromeda galaxy. Which spectrum is from Andromeda?

Spectra for the sun and two galaxies. Source: Karla Panchuk (2015) CC BY 4.0.

3. Astronomers looking for some of the earliest stars in the universe were surprised to find a planetary system called HIP 11952, which existed 12.8 billion years ago. This was very early in the universe's history, when stars still consisted largely of hydrogen and helium. Do you think there were terrestrial planets in this system? Why or why not?

4. Summarize the trends in size and composition of objects in the solar system.

5. What is the frost line, and what does it help to explain?

6. Why is Pluto not considered a planet?

7. What is differentiation?

8. The exoplanet Kepler-452b is within the habitable zone of its star. In our solar system, planets a similar distance from the Sun are terrestrial planets. Why can we not say for certain that Kepler-452b's distance from its star means it is a terrestrial planet?

9. Of the planetary systems discovered thus far, none are exactly like our solar system. Does this mean our solar system is unique in the universe?

Answers at the end of the chapter

References

European Space Agency (2015). *Planck reveals first stars were born late*. Retrieved from http://www.esa.int/Our_Activities/Space_Science/Planck/Planck_reveals_first_stars_were_born_late

Extrasolar Planets Catalog. Available from http://exoplanet.eu/catalog/

Habitable Exoplanets Catalog. Available from http://phl.upr.edu/projects/habitable-exoplanets-catalog

Knop, R. (2010). *The History of the Universe*.. Retrieved from http://galacticinteractions.scientopia.org/2010/08/02/the-history-of-the-universe/

Lawrence, C. R. (2015, March). Planck 2015 Results. Paper presented to the Astrophysics Subcommittee, NASA HQ. Retrieved from https://smd-prod.s3.amazonaws.com/science-green/s3fs-public/mnt/medialibrary/2015/04/08/CRL_APS_2015-03-18_compressed2.pdf

Pilipenko, S. V. (2013). Paper-and-pencil cosmological calculator. arXiv:1303.5961v1 [astro-ph.CO]

Royal Astronomical Society (2016, June 28). CR7 is not alone—A team of super bright galaxies in the early universe. *Phys.org* Retrieved from https://phys.org/news/2016-06-cr7-alonea-team-super-bright.html

Sobral, D., Matthee, J., Darvish, B., Schaerer, D., Mobasher, B., Röttgering, H., Santos, S., & Hemmati, S. (2015). Evidence for PopIII-like stellar populations in the most luminous Lyman-α emitters at the epoch of re-ionisation: spectroscopic confirmation. *The Astrophysical Journal 808*(2) doi: 10.1088/0004-637x/808/2/139.

Answers to Review Questions

1. To see an event, light from that event must reach our eyes. Light travels very quickly (about 300,000,000 m/s), but the universe is very, very large. Depending on how far away the event was, it could take billions of years for light to travel from the event to our eyes so we can see it. Astronomers take advantage of this fact to view the universe's past.

2. B is the spectrum from the Andromeda galaxy. We know that one spectrum represents the sun, which is not moving toward or away from us. (Our orbit is not perfectly circular, but the small eccentricity is not a factor in this comparison.) We know that the Andromeda galaxy is on a collision course with us, so it is the exception to the rule that galaxies are moving away from us, and their light is red-shifted. That means the spectrum B which is shifted furthest to the left (blue-shifted) is Andromeda, and spectrum A which is furthest to the right (red-shifted) is a galaxy moving away from us. That means C is the sun.

3. The planetary system consisted of two Jupiter-sized gas giant planets. Gas giant planets contain large amounts of hydrogen, and hydrogen was plentiful in the early universe. In contrast, terrestrial planets have heavier elements, especially silica, iron, magnesium, and nickel, that had yet to be manufactured by stars. Those elements were not present in sufficient abundance to form terrestrial planets until much later.

4. Closest to the sun we find the small, rocky, terrestrial planets with metal cores. Further out are the gas giant planets, which are the largest in the solar system. They consist mostly of hydrogen, and have cores of rock and ice. Beyond the gas giant planets are the ice giant planets, which are next largest. They have a mantle of ice (not just water ice but ammonia and methane ice), and a rocky core. Smaller objects in the solar system include rocky bodies within the asteroid belt between Mars and Jupiter, and bodies of ice and dust in the Kuiper belt and Oort cloud beyond Neptune.

5. The frost line marks the distance from the sun beyond which temperatures were cool enough to allow ice to form. This helps to explain why the terrestrial planets are closer to the sun, and the Jovian and ice giant planets farther away. Mineral grains could solidify and begin to accrete closer to the sun, forming terrestrial planets, because they have higher melting points. In contrast, water vapour, methane, and ammonia had to be farther from the sun before they could freeze and begin to accrete.

6. Planets are defined as having cleared their orbits of debris. Pluto is located within the Kuiper belt, so it shares its orbit with other objects. There are two other criteria in the definition of a planet: planets in our solar system must orbit the sun, and they must have a spherical shape. Pluto satisfies both these criteria, but sadly the people deciding whether or not Pluto should be a planet are not amenable to a "best two out of three" compromise.

7. Differentiation is the separation of materials within a planet such that dense materials sink to the core. In Earth's case, the denser materials are iron and nickel.

8. The fact that we have terrestrial planets close to the sun makes sense in terms of the frost line, but it does not seem to be a hard-and-fast rule in other planetary systems. Therefore, we can't conclude from Kepler-452b's position alone that it is a terrestrial planet.

9. The rules of the accretion game mean that there are many complex interactions, so even a small difference in the starting conditions or in how the game goes in the beginning could have major implications in the end. For that reason, we shouldn't expect to find a planetary system that matches ours in every minute detail. However, just because we haven't found a similar planetary system does not mean one does not exist. Our planet-finding methods are biased toward discovering large planets orbiting close to their stars, whereas our solar system has small planets close to the sun and larger ones farther away. That doesn't mean our methods won't eventually turn up a system like ours, just that they are more likely to turn up systems that are different.

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-NC-SA 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 3. Earth's Interior

Adapted by Karla Panchuk, University of Saskatchewan from *Physical Geology* by Steven Earle

Figure 3.1 The red rocks of the Tablelands in Gros Morne National Park are a sample of Earth's mantle. Top: The red rocks of the Tablelands are on the right, and stand in contrast with the green surroundings. Bottom: A closer view of Tablelands terrain, showing rocks weathered red, and a near absence of plant life. *Source: Top photograph- Leos Kral (2008) CC BY-NC-SA 2.0; Bottom photograph: Tara Joyce (2013) CC BY-SA 2.0. See Appendix C for more attributions.*

Learning Objectives

After reading this chapter and answering the review questions at the end, you should be able to:

- Explain the variations in the composition and characteristics of Earth's different layers.
- Explain how seismic data can be used to understand the structure of Earth's interior.
- Describe the temperature variations within Earth and their implications for internal processes such as mantle convection.
- Explain the origins of Earth's magnetic field and the timing of magnetic field reversals.
- Describe the isostatic relationship between the crust and the mantle, and the implications of that relationship for geological processes on Earth

The barren red rocks of the Tablelands stand in stark contrast to their lush green surroundings in Gros Morne National Park (Figure 3.1, top). If the Tablelands appear out of place, it's because they are. The Tablelands are one of few places on Earth where you can walk directly on the rocks of Earth's mantle, thanks to an accident of plate tectonics that happened hundreds of millions of years ago. The red colour of

the Tablelands rocks comes from iron-bearing minerals reacting with oxygen. Unaltered, the rocks are dark green (Figure 3.2). The rocks lack vegetation because the chemical composition of the rocks does not provide adequate nutrients for plants.

Figure 3.2 Tablelands mantle rock with reddish weathering rind, and dark green fresh surface. Scale in cm. *Source: Karla Panchuk (2017) CC BY 4.0*

Locations like the Tablelands are one way we can learn about Earth's interior. Meteorites derived from smashed differentiated bodies (asteroids that separated into mantle and core) are another. Asteroids that formed at a similar distance from the sun as Earth had a mineral composition akin to Earth's. When these objects were shattered in giant collisions, the result was stony meteorites from fragmented mantle rock, and iron meteorites from fragmented core. Some fragments sampled the result of violent encounters that mixed the two (Figure 3.3).

Figure 3.3 Cut and polished slab of a stony-iron meteorite called a pallasite, thought to have formed in a collision that smashed mantle rocks against the metal core of an asteroid early in the solar system's history. Green and brown crystals are the mineral olivine. The metal between the olivine crystals is an iron-nickel mineral. *Source: Muséum de Toulouse (2012) CC BY-NC 2.0*

We also get information about the structure of Earth's interior by analyzing the speeds and paths of earthquake vibrations, called **seismic waves**.

We need to know something about the inside of our planet— what it's made of, and what happens within it— in order to understand how Earth works, especially the mechanisms of plate tectonics. It is fortunate that there are many ways for geologists gather information about Earth's interior, because one thing they can't do is go down and look at it.

3.1 Earth's Layers: Crust, Mantle, & Core

Earth consists of three main layers: the **crust**, the **mantle**, and the **core** (Figure 3.4). The core accounts for almost half of Earth's radius, but it amounts to only 16.1% of Earth's volume. Most of Earth's volume (82.5%) is its mantle, and only a small fraction (1.4%) is its crust.

Figure 3.4 Earth's interior. Right- crust, mantle, and outer and inner core to scale. Left- Cutaway showing continental and ocean crust, and upper mantle layers. The lithosphere is the crust plus the uppermost layer of the mantle. *Source: Karla Panchuk (2018) CC BY 4.0. Earth photo by NASA (n.d.) Public Domain*

Crust

The Earth's outermost layer, its **crust**, is rocky and rigid. There are two kinds of crust: **continental crust**, and **ocean crust**. Continental crust is thicker, and predominantly **felsic** in composition, meaning that it contains minerals that are richer in silica. The composition is important because it makes continental crust less dense than ocean crust.

Ocean crust is thinner, and predominantly **mafic** in composition. Mafic rocks contain minerals with less silica, but more iron and magnesium. Mafic rocks (and therefore ocean crust) are denser than the felsic rocks of continental crust.

The crust floats on the mantle. Continental crust floats higher in the mantle than ocean crust because of the lower density of continental crust. An important consequence of the difference in density is that if tectonic plates happen to bring ocean crust and continental crust into collision, the plate with ocean crust will be forced down into the mantle beneath the plate with continental crust.

Mantle

The mantle is almost entirely solid rock, but it is in constant motion, flowing very slowly. It is **ultramafic** in composition, meaning it has even more iron and magnesium than mafic rocks, and even less silica. Although the mantle has a similar chemical composition throughout, it has layers with different

mineral compositions and different physical properties. It can have different mineral compositions and still be the same in chemical composition because the increasing pressure deeper in the mantle causes mineral structures to be reconfigured.

Rocks higher in the mantle are typically composed of **peridotite**, a rock dominated by the minerals olivine and pyroxene. The Tablelands rock in Figure 3.2 is a type of peridotite. Lower in the mantle, extreme pressures transform minerals and create rocks like **eclogite** (Figure 3.5), which contains garnets.

Figure 3.5 Eclogite from the Swiss-Italian Alps. Reddish brown spots are garnets. *Source: James St. John (2014) CC BY 2.0*

Lithosphere

The lithosphere can't be classified neatly as either crust or mantle because it consists of both. It is formed from the crust as well as the uppermost layer of the mantle which is stuck to the underside of the crust. Tectonic plates are fragments of lithosphere.

Asthenosphere

Beneath the lithosphere is the asthenosphere.Tiny amounts of melted rock dispersed through the otherwise solid asthenosphere make the asthenosphere weak compared to the lithosphere. The weakness of the asthenosphere is important for plate tectonics because it deforms as fragments of lithosphere move around upon and through it. Without a weak asthenosphere, plates would be locked in place, unable to move as they do now.

D"

The D" (dee double prime) layer is a mysterious layer beginning approximately 200 km above the boundary between the core and mantle. (This boundary is referred to as the core-mantle boundary.) We know it exists because of how seismic waves change speed as they move through it, but it isn't clear why it's different from the rest of the mantle. One idea is that it is minerals are undergoing another transition in this region because of pressure and temperature conditions, similar to the transition between the upper and lower mantle. Other ideas are that small pools of melt are present, or that the differences in seismic properties are due to subducted slabs of lithosphere resting on the core-mantle boundary.

Core

The core is primarily composed of iron, with lesser amounts of nickel. Lighter elements such as sulfur, oxygen, or silicon may also be present. The core is extremely hot (~3500° to more than 6000°C). But despite the fact that the boundary between the inner and outer core is approximately as hot as the surface of the sun, only the outer core is liquid. The inner core is solid because the pressure at that depth is so high that it keeps the core from melting.

3.2 Imaging Earth's Interior

Seismology is the study of vibrations within Earth. These vibrations are caused by events such as earthquakes, extraterrestrial impacts, explosions, storm waves hitting the shore, and tides. Seismology is applied to the detection and study of earthquakes, but seismic waves also provide important information about Earth's interior.

Seismic waves travel through different materials at different speeds, and we can apply knowledge of how they interact with different materials to understand Earth's layers and internal structures. Similar to the way that ultrasound is used to image the human body, we can measure how long it takes for seismic waves to travel from their source to a recording station.

Another feature of seismic waves is that some, called **P-waves**, can travel rapidly though both liquids and solids, but others, called **S-waves**, can only travel though solids, and are slower than P-waves. Observing where P-waves travel, and S-waves do not, allows us to identify regions within Earth that are melted.

Seismic Wave Paths

Seismic waves travel in all directions from their source, but it is convenient to imagine the path traced by one point on the wave front, and represent that path as a **seismic ray** (arrows, Figure 3.6).

Figure 3.6 Seismic waves and seismic rays. The paths of seismic waves can be represented as rays. Seismic ray paths are bent when they enter a rock layer with a different seismic velocity. *Source: Karla Panchuk (2018) CC BY 4.0*

When seismic waves encounter a different rock layer, some might bounce off the layer, or **reflect**, as in the bottom layer of Figure 3.6. But some waves will travel through the layer. If the wave travels at a different speed in the new layer, its path will be bent, or **refracted**, as it crosses into the new layer. If the wave can travel faster in the new layer, it will be bent slightly toward the contact between the two layers. In Figure 3.6, the ray can travel progressively faster in each layer as it goes down through the layers, and it is bent slightly upward each time it crosses into the next layer. The reverse happens if the wave slows down. On the right side of the diagram, the wave is moving upward through slower and slower layers. It is bent away from the faster layer each time, causing it to take a more direct path to the surface.

Seismic velocities are higher in more rigid layers, so broadly speaking, they get faster deeper within Earth, because higher pressures make layers more rigid. They tend to take curved paths through the Earth because refraction bends their path until they are reflected and directed upward again, as in Figure 3.6.

Discoveries with Seismic Waves

The Moho: Where Crust Meets Mantle

One of the first discoveries about Earth's interior made through seismology was in the early 1900s by Croatian seismologist Andrija Mohorovičić (pronounced *Moho-ro-vi-chich*). He noticed that sometimes, seismic waves arrived at seismic stations (measuring locations) farther from an earthquake *before* they arrived at closer ones. He reasoned that the waves that traveled farther were faster because they bent down and traveled faster through different rocks (those of the mantle) before being bent upward back into the crust (Figure 3.7).

Figure 3.7 Depiction of seismic waves emanating from an earthquake (red star). Some waves travel through the crust to the seismic station (at ~6 km/s), while others go down into the mantle (where they travel at ~8 km/s) and are bent upward toward the surface, reaching the station before the ones that travelled only through the crust. *Source: Steven Earle (2016) CC BY 4.0*

The boundary between the crust and the mantle is now known as the **Mohorovičić discontinuity** (or Moho). Its depth is between 60 – 80 km beneath major mountain ranges, 30 – 50 km beneath most of the continental crust, and 5 – 10 km beneath ocean crust.

The Core-Mantle Boundary

Arguments for a liquid outer core were supported by a distinctive signature in the global distribution of seismic waves from earthquakes. When an earthquake occurs, there is a zone on the opposite side of Earth where S-waves are not measured. This **S-wave shadow zone** begins 103° on either side of the earthquake, for a total angular distance of 154° (Figure 3.8, left). There is also a **P-wave shadow zone** on either side of the earthquake, from 103° to 150° (Figure 3.8, right).

Figure 3.8 Patterns of seismic wave propagation through Earth's mantle and core. S-waves do not travel through the liquid outer core, so they leave a shadow on Earth's far side. P-waves do travel through the core, but because the waves that enter the core are refracted, there are also P-wave shadow zones. *Source: Steven Earle (2016) CC BY 4.0*

The S-wave shadow zone occurs because S-waves cannot travel through the liquid outer core. The P-wave shadow zone occurs because seismic velocities are much lower in the liquid outer core than in the overlying mantle, and the P-waves are refracted in a way that leaves a gap. Not only do the shadow zones tell us that the outer core is liquid, the size of the shadow zones allows us to calculate the size of the core, and the location of the core-mantle boundary.

Seismic Portrait of Earth's Layers

The change seismic wave velocity with depth in Earth (Figure 3.9) has been determined over the past several decades by analyzing seismic signals from large earthquakes all around the world. Earth's layers are detectable as changes in velocity with depth. The asthenosphere is visible as a low velocity zone within the upper mantle (Figure 3.9, left). There is an abrupt increase in P-wave velocity at 420 km, showing the depth at which minerals transform into structures that are more stable at higher pressures and temperatures.

Figure 3.9 P-wave and S-wave velocity variations with depth from the crust through the upper mantle (left) and from the crust through to the core (right). *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2016) CC BY 4.0*

The boundary between the upper and lower mantle is visible at 660 km as a sudden change from rapidly increasing P- and S-wave velocities to slow or no change in P-wave and S-wave velocities (Figure 3.9, right). The core-mantle boundary (CMB in Figure 3.9) is apparent as a sudden drop in P-wave velocities, where seismic waves move from solid mantle to liquid outer core. The boundary between the outer core and inner core is marked by a sudden increase in P-wave velocity after 5000 km, where seismic waves move from a liquid back into a solid again.

Seismic Images of Plate Tectonic Structures

Using data from many seismometers and hundreds of earthquakes, it is possible to create images from the seismic properties of the mantle. This technique is known as **seismic tomography**. Tomography can be used to map out slabs of lithosphere that are entering the mantle, or have disappeared within it. Those slabs

are cooler, and therefore more rigid than surrounding mantle rocks, so seismic waves travel through them faster. In Figure 3.10, higher-than-average seismic velocities in cool slabs are indicated in dark blue.

Figure 3.10 P-waves and S-waves used to map out the location of the Cocos slab of lithosphere. The slab appears in dark blue, indicating higher than average seismic wave velocities. Left- Tomograms showing seismic wave anomalies for a 1290 km surface. Right- Cross-sections along the transect marked X-Y on the globe. *Source: Karla Panchuk (2018) CC BY 4.0, modified after van der Meer et al. (2018) CC BY 4.0*

Thanks to the tomograms, we can see that the Cocos plate, which is colliding with Central America, is part of a much larger slab of lithosphere that has already settled onto the mantle. Tomograms representing a surface at 1290 km depth (Figure 3.10, left) show that at that level, the Cocos slab is beneath the Caribbean Sea. The tomograms on the right show a vertical view along the line X-Y marked on the globe. The vertical tomograms show us that the Cocos slab extends all the way down to the core-mantle boundary.

Visit the Underworld

What is the Atlas of the Underworld?

The *Atlas of the Underworld* is a catalog of more than 90 slabs of lithosphere that have been imaged within the mantle using seismic tomography. The *Atlas* includes tomographic images, locator maps, and geological histories for each slab. The catalog can be searched online at http://www.atlas-of-the-underworld.org/ or viewed in the original publication by van der Meer et al. (2018). The *Atlas of the Underworld* is an open-access resource.

The HADES Underworld Explorer

Create your own tomographic cross-sections for locations anywhere in the world by using this intuitive drag-and-drop tool. Visit the HADES Underworld Explorer at http://www.atlas-of-the-underworld.org/hades-underworld-explorer/

3.3 Earth's Interior Heat

Earth Gets Hotter the Deeper You Go

Earth's temperature increases with depth, but not at a uniform rate (Figure 3.11). Earth's **geothermal gradient** is 15° to 30°C/km within the crust. It then drops off dramatically through the mantle, increases more quickly at the base of the mantle, and then increases slowly through the core. The temperature is approximately 1000°C at the base of the crust, around 3500°C at the base of the mantle, and approximately 6,000°C at Earth's centre.

Figure 3.11 Geothermal gradient (change in temperature with depth). Left- Geothermal gradient in the crust and upper mantle. The geothermal gradient remains below the melting temperature of rock, except in the asthenosphere. There, temperatures are high enough to melt some of the minerals. Right- Geothermal gradient throughout Earth. Rapid changes occur in the uppermost mantle, and at the core-mantle boundary. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2016) CC BY 4.0*

In spite of high temperatures within Earth, mantle rocks are almost entirely solid. High pressures keep them from melting. The red dashed line in Figure 3.11 (right) shows the minimum temperature at which dry mantle rocks will melt. Rocks at temperatures to the left of the line will remain solid. In rocks at temperatures to the right of the line, *some* minerals will begin to melt. Notice that the red dashed line goes further to the right for greater depths, and therefore greater pressures. Now compare the geothermal gradient with the red dashed line. The geothermal gradient is to the left of the red line, except in the asthenosphere, where small amounts of melt are present.

Convection Helps to Move Heat Within Earth

The fact that the temperature gradient is much lower in the main part of the mantle than in the lithosphere has been interpreted as evidence of **convection** in the mantle. When the mantle convects, heat is transferred through the mantle by physically moving hot rocks. Mantle convection is the result of heat transfer from the core to the base of the lower mantle. As with a pot of soup on a hot stove (Figure 3.12), the material near the heat source (the soup at the bottom of the pot) becomes hot and expands, making it less dense than

the material above. Buoyancy causes it to rise, and cooler material flows in from the sides. Of course, convection in the soup pot is much faster than convection in the mantle. Mantle convection occurs at rates of centimetres per year.

Figure 3.12 Convection in a pot of soup on a hot stove (left). As long as heat is being transferred from below, the liquid will convect. If the heat is turned off (right), the liquid remains hot for a while, but convection will cease. *Source: Steven Earle (2015) CC-BY 4.0*

Convection carries heat to the surface of the mantle much faster than heating by **conduction**. Conduction is heat transfer by collisions between molecules, and is how heat is transferred from the stove to the soup pot. A convecting mantle is an essential feature of plate tectonics, because the higher rate of heat transfer is necessary to keep the asthenosphere weak. Earth's mantle will stop convecting once the core has cooled to the point where there is not enough heat transfer to overcome the strength of the rock. This has already happened on smaller planets like Mercury and Mars, as well as on Earth's moon. When mantle convection stops, the end of plate tectonics will follow.

Models of Mantle Convection

In the soup pot example, convection moves hot soup from the bottom of the pot to the top. Some geologists think that Earth's convection works the same way— hot rock from the base of the mantle moves all the way to the top of the mantle before cooling and sinking back down again. This view is referred to as **whole-mantle convection** (Figure 3.8, left). Other geologists think that the upper and lower mantle are too different to convect as one. They point to slabs of lithosphere that are sinking back into the mantle, some of which seem to perch on the boundary between the upper and lower mantle, rather than sinking straight through. They also note chemical differences in magma originating in different parts of the mantle— differences that are not consistent with the entire mantle being well stirred. They argue that **double-layered convection** is a better fit with the observations (Figure 3.13, right). Still others argue that there may be some locations where convection goes from the bottom of the mantle to the top, and some where it doesn't (Figure 3.13, middle).

Figure 3.13 Models of mantle convection. Left- whole mantle convection. Rocks rise from the core-mantle boundary to the top of the mantle, then sink to the bottom again. Right- Two-layer convection, in which upper and lower mantle convect at different rates. Middle- Convection paths vary depending on the circumstances. *Source: Karla Panchuk (2018) CC BY 4.0*

Why Is Earth Hot Inside?

The heat of Earth's interior comes from a variety of sources. These include the heat contained in the objects that accreted to form Earth, and the heat produced when they collided. As Earth grew larger, the increased pressure on Earth's interior caused it to compress and heat up. Heat also came from friction when melted material was redistributed within Earth, forming the core and mantle.

A major source of Earth's heat is **radioactivity**, the energy released when the unstable atoms decay. The radioactive isotopes uranium-235 (^{235}U), uranium-238 (^{238}U), potassium-40 (^{40}K), and thorium-232 (^{232}Th) in Earth's mantle are the primary source. Radioactive decay produced more heat early in Earth's history than it does today, because fewer atoms of those isotopes are left today (Figure 3.14). Heat contributed by radioactivity is now roughly a quarter what it was when Earth formed.

Figure 3.14 Production of heat within the Earth over time by radioactive decay of uranium, thorium, and potassium. Heat production has decreased over time as the abundance of radioactive atoms has decreased. *Source: Steven Earle (2015) CC BY 4.0, modified after Arevalo et al. (2009)*

3.4 Earth's Magnetic Field

Earth's liquid iron core convects because it is heated from beneath by the inner core. Because iron is a metal and conducts electricity (even when molten), its motion generates a magnetic field.

Earth's magnetic field is defined by north and south poles representing lines of magnetic force flowing into Earth in the northern hemisphere and out of Earth in the southern hemisphere (Figure 3.15). Because of the shape of the field lines, the magnetic force is oriented at different angles to the surface in different locations. The tilt, or **inclination** of magnetic field lines is represented by the tilt of compass needles in Figure 3.15. At the north and south poles, the force is vertical. The force is horizontal at the equator. Everywhere in between, the magnetic force is at an intermediate angle to the surface.

Figure 3.15 Earth's magnetic field depicted as the field of a bar magnet coinciding with the core. The south pole of the magnet points to Earth's magnetic north pole. The red and white compass needles represent the orientation of the magnetic field at various locations on Earth's surface. *Source: Karla Panchuk (2018) CC BY-SA 4.0, modified after Steven Earle (2015) CC BY-SA 4.0, and T. Stein (2008) CC BY-SA 3.0 view source*

Exercise: Magnetic Inclination

Regular compasses point only to the north magnetic pole, but if you had a magnetic dip meter (or a smartphone with the appropriate app), you could also measure the angle of the magnetic field at your location in the up-and-down sense. However, you don't need a dip meter or app to do this exercise! Using Figure 3.15 as a guide, describe the general location on Earth where the vertical angles would be as follows:

1. Straight down
2. Down at a steep angle
3. Up at a steep angle
4. Parallel to flat ground

Earth's magnetic field is generated within the outer core by the convective movement of liquid iron, but although convection is continuous, the magnetic field is not stable. Periodically, the magnetic field decays and then becomes re-established. When it does re-establish, the polarity may have reversed (i.e., your compass would point south rather than north). Over the past 250 Ma, there have been hundreds of magnetic field reversals, and their timing has been anything but regular. The shortest ones that geologists have been able to identify lasted only a few thousand years, and the longest one was more than 30 million years, during the Cretaceous Period (Figure 3.16).

Figure 3.16 Magnetic field reversal chronology for the past 170 Ma. Black stripes mark times when the magnetic field was oriented the same as today. *Source: Steven Earle (2015) CC BY 4.0, modified after AnomieX (2010) Public Domain*

Changes in Earth's magnetic field have been studied using mathematical models that simulate convection in the outer core (Figure 3.17). Reversals happened spontaneously when the model was run to simulate a period of several hundred thousand years. Spontaneous reversals can happen because convection does not occur in an orderly way, in spite of what the bar magnet analogy may suggest. Many small-scale variations occur in convection patterns within the inner core, and Earth's magnetic field over all is the sum of those variations. Magnetic reversals do not happen as frequently as they might, if not for the solid inner core. Magnetic field changes take much longer within the inner core, so reversals in the outer core do not always coincide with reversals in the inner core. Both are required in order for Earth's magnetic field to flip.

Figure 3.17 Earth's magnetic field between reversals (left) and during a reversal (right). The lines represent magnetic field lines: blue where the field points toward Earth's centre and yellow where it points away. The rotation axis of Earth is vertical, and the outline of the core is shown as a dashed white circle. *Source: NASA (2007) Public Domain*

3.5 Isostasy

Lithospheric Plates Float on the Mantle

The mantle is able to convect because it can deform by flowing over very long timescales. This means that tectonic plates are *floating* on the mantle, like a raft floating in the water, rather than *resting* on the mantle like a raft sitting on the ground. How high the lithosphere floats will depend on the balance between gravity pulling the lithosphere down, and the force of buoyancy as the mantle resists the downward motion of the lithosphere. **Isostasy** is the state in which the force of gravity pulling the plate toward Earth's centre is balanced by the resistance of the mantle to letting the plate sink.

To see how isostasy works, consider the rafts in Figure 3.18. The raft on the right is sitting on solid concrete. The raft will remain at the same elevation whether there are two people on it, or four, because the concrete is too strong to deform. In contrast, isostasy is in play for the rafts on the left, which are floating in a swimming pool full of peanut butter. With only one person on board, the raft floats high in the peanut butter, but with three people, it sinks dangerously low. Peanut butter, rather than water, is used in this example because the viscosity of peanut butter (its stiffness or resistance to flowing) more closely represents the relationship between the tectonic plates and the mantle. Although peanut butter has a similar density to water, it's higher viscosity means that if a person is added to a raft, it will take longer for the raft to settle lower into the peanut butter that it would take the raft to sink into water.

Figure 3.18 Illustration of isostatic relationships between rafts and peanut butter (left), and a non-isostatic relationship between a raft and solid ground (right). *Source: Steven Earle (2015) CC BY 4.0*

The relationship of Earth's crust to the mantle is similar to the relationship of the rafts to the peanut butter. The raft with one person on it floats comfortably high. Even with three people on it the raft is less dense than the peanut butter, so it floats, but it floats uncomfortably low for those three people. The crust, with an average density of around 2.6 g/cm^3, is less dense than the mantle (average density of ~3.4 g/cm^3 near the surface, but more at depth), and so it is floating on the mantle. When weight is added to the crust through the process of mountain building, the crust slowly sinks deeper into the mantle, and the mantle material that was there is pushed aside (Figure 3.19, left). When erosion removes material from the mountains over tens of millions of years, decreasing the weight, the crust rebounds and the mantle rock flows back (Figure 3.19, right).

Figure 3.19 Illustration of isostatic relationships between rafts and peanut butter (left), and a non-isostatic relationship between a raft and solid ground (right). *Source: Steven Earle (2015) CC BY 4.0*

Isostasy and Glacial Rebound

The crust and mantle respond in a similar way to glaciation. Thick accumulations of glacial ice add weight to the crust, and the crust subsides, pushing the mantle out of the way. The Greenland ice sheet, at over 2,500 m thick, has depressed the crust below sea level (Figure 3.20a). When the ice eventually melts, the crust and mantle will slowly rebound (Figure 3.20b), but full rebound will likely take more than 10,000 years (3.20c).

Large parts of Canada are still rebounding as a result of the loss of glacial ice over the past 12,000 years, as are other parts of the world (Figure 3.21). The highest rate of uplift is in a large area west of Hudson Bay, where the Laurentide Ice Sheet was the thickest, at over 3,000 m. Ice finally left this region around 8,000 years ago, and the crust is currently rebounding at nearly 2 cm/year. Strong isostatic rebound is also occurring in northern Europe where the Fenno-Scandian Ice Sheet was thickest, and in the eastern part of Antarctica, which also experienced significant ice loss during the Holocene.

Glacial rebound in one location means subsidence in surrounding areas (Figure 3.21, yellow through red regions). Regions surrounding the former Laurentide and Fenno-Scandian Ice Sheets that were lifted up when mantle rock was forced aside and beneath them are now subsiding as the mantle rock flows back.

Figure 3.20 Cross-section through the crust in the northern part of Greenland. a) Up to 2,500 m of ice depresses the crust downward (red arrows) and below sea level. b) After complete melting. Isostatic rebound would be slower than the rate of melting, leaving central Greenland at or below sea level for thousands of years. c) Complete rebound after ~10,000 years raises central Greenland above sea level again. *Source: Steven Earle (2015) CC BY 4.0*

Figure 3.21 Current rates of post-glacial isostatic uplift (green, blue, and purple shades) and subsidence (yellow and orange). Subsidence is taking place where the mantle is slowly flowing back toward areas that are experiencing post-glacial uplift. *Source: Steven Earle (2015) CC BY 4.0, modified after Erik Ivins, JPL (2010) Public Domain*

How Can the Mantle Be Both Solid and Plastic?

You might be wondering how it is possible that Earth's mantle is solid, rigid rock, and yet it convects and flows like a very viscous liquid. The explanation is that the mantle behaves as a non-Newtonian fluid, meaning that it responds differently to stresses depending on how quickly the stress is applied.

A good example of non-Newtonian behaviour is the deformation of Silly Putty, which can bounce when it is compressed rapidly when dropped, and will break if you pull on it sharply. But will deform in a liquid manner if stress is applied slowly. The force of gravity applied over a period of hours can cause it to deform like a liquid, dripping through a hole in a glass tabletop (Figure 3.22). Similarly, the mantle will flow when placed under the slow but steady stress of a growing (or melting) ice sheet.

Chapter 3 Summary

Figure 3.22 Silly Putty exhibiting plastic behaviour when acted upon by gravity over several hours. *Source: Erik Skiff (2006) CC BY-SA 2006*

The topics covered in this chapter can be summarized as follows:

3.1 Earth's Layers

Earth is divided into a rocky crust and mantle, and a core consisting largely of iron. The crust and the uppermost mantle form the lithosphere, which is broken into tectonic plates. The next layer, the asthenosphere, allows the plates to move because it deforms by flowing.

3.2 Imaging Earth's Interior

Seismic waves that travel through Earth are either P-waves or S-waves. P-waves are faster than S-waves, and can pass through fluids. Earth's layers can be identified by looking at changes in the velocity of seismic waves. Seismic wave shadow zones contributed to knowledge of the depth of the core-mantle boundary, and the knowledge that the outer core is liquid. Plate tectonic structures within Earth can also be mapped using the seismic waves generated by earthquakes.

3.3 Earth's Interior Heat

Earth's temperature increases with depth (to around 6000°C at the centre), but the rate of increase is not the same everywhere. In the lithosphere, thickness and plate tectonic setting are are factors. Deeper within the mantle, convection currents are more important.

3.4 Earth's Magnetic Field

Earth's magnetic field is generated by convection of the liquid outer core. The magnetic field is similar to that of a bar magnet, and has force directions that vary with latitude. The polarity of the field is not constant, meaning that the positions of the north and south magnetic poles have flipped from "normal" (as it is now) to reversed and back many times in Earth's history.

3.5 Isostasy

The plastic nature of the mantle, which allows for mantle convection, also determines the nature of the relationship between the crust and the mantle. The crust floats on the mantle in an isostatic relationship. Where the crust becomes thicker and heavier because of mountain building, it pushes farther down into the mantle. Oceanic crust, being denser than continental crust, floats lower on the mantle than continental crust.

Review Questions

1. What parts of Earth are most closely represented by stony meteorites and iron meteorites?
2. Draw a simple diagram of Earth's layers, and label the approximate locations of the following boundaries: crust/mantle, mantle/core, outer core/inner core.
3. How do P-waves and S-waves differ?

4. Why does P-wave velocity decrease suddenly at the core-mantle boundary?
5. Why do both P-waves and S-waves gradually bend as they move through the mantle?
6. What is the evidence for mantle convection, and what causes mantle convection?
7. How is Earth's magnetic field generated?
8. When were the last two reversals of Earth's magnetic field?
9. What property of the mantle is essential for the isostatic relationship between the crust and the mantle?
10. How would you expect the depth to the crust-mantle boundary in the area of the Rocky Mountains to differ from that in central Saskatchewan?
11. British Columbia is still experiencing weak post-glacial isostatic uplift, especially in the interior, but also along the coast (see Figure 3.21). Meanwhile, offshore areas are experiencing weak isostatic subsidence. Why?

Answers at the end of the chapter

References

Arevalo, R., McDonough, W., & Luong, M. (2009). The K/U ratio of Earth: Insights into mantle composition, structure and thermal evolution. *Earth and Planetary Science Letters, 278*(3-4), 361-369. https://doi.org/10.1016/j.epsl.2008.12.023

British Geological Survey, Natural Environment Research Council (n.d.). *Reversals: Magnetic Flip*. Retrieved from http://www.geomag.bgs.ac.uk/education/reversals.html

Glatzmaier, G. A. (n.d.) *The Geodynamo*. Retrieved from https://websites.pmc.ucsc.edu/~glatz/geodynamo.html

van der Meer, D.G., van Hinsbergen, D.J.J., and Spakman, W., (2018). Atlas of the Underworld: slab remnants in the mantle, their sinking history, and a new outlook on lower mantle viscosity. *Tectonophysics 723*, p. 309-448. https://doi.org/10.1016/j.tecto.2017.10.004

Answers to Review Questions

1. Stony meteorites are similar in composition to Earth's mantle, while iron meteorites are similar to the core.

2. Compare your answer to Figure 3.4.

3. P-waves can pass through a liquid, and travel approximately twice as fast as S-waves (which cannot pass through a liquid).

4. P-wave velocity decreases at the core-mantle boundary because the outer core is liquid.

5. The mantle gets increasingly dense and strong with depth because of the increasing pressure. This difference affects both P-wave and S-wave velocities, and they are refracted toward the lower density mantle material (meaning they are bent out toward Earth's surface).

6. The key evidence for mantle convection is that the rate of temperature increase with depth within the mantle is less than expected. This can only be explained by a mantle that is mixing by convection. The mechanism for convection is the transfer of heat from the core to the mantle, causing the to mantle flow.

7. Earth's magnetic field is generated within the liquid outer core because liquid metal is convecting.

8. The last two reversals of Earth's magnetic field were at the beginning of the present Brunhes normal chron (0.78 Ma), and at the end of the Jaramillo normal subchron (0.90 Ma).

9. The isostatic relationship between the crust and the mantle is dependent on the fact that over very long timescales, the mantle deforms by flowing.

10. In the area of the Rocky Mountains the crust is thickened and pushed down into the mantle. In Saskatchewan the crust is thinner and does not extend as far into the mantle.

11. During the Pleistocene glaciation, British Columbia was pushed down by glacial ice. Mantle rock flowed slowly out from under the weighted-down crust and toward the ocean floor. Now that the land area is rebounding, that mantle rock is flowing back and the offshore areas are subsiding.

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-NC-SA 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 4. Plate Tectonics

Adapted by Karla Panchuk from *Physical Geology* by Steven Earle

Figure 4.1 Iceland is known for its volcanoes, which are present because Iceland is located on the Mid-Atlantic Ridge, where the Atlantic Ocean is spreading apart and new crust is forming. In fact, Iceland exists because that volcanic activity has built up the island from the ocean floor. Iceland is cut by rift zones (white lines on the map at left) where the island is splitting apart along with the rest of the Atlantic Ocean. Rift zones are marked by belts of young volcanic rocks (dark green). You can stand on a rift zone if you visit Thingvellir National Park (right). Rifting has produced a valley where the crust has settled downward. The margins of the North American and Eurasian tectonic plates are visible as ridges on either side of the valley. The photographer was standing on a ridge on the North American side. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Photo: Ruth Hartnup (2005) CC BY 2.0. See Appendix C for more attributions.*

Learning Objectives

After reading this chapter and answering the review questions at the end, you should be able to:

- Discuss the early evidence for continental drift, and Alfred Wegener's role in promoting this theory.
- Describe other models that were used early in the 20th century to understand global geological features.
- Summarize the geological advances that provided the basis for understanding the mechanisms of plate tectonics, and the evidence that plates and are constantly being created and destroyed.
- Describe the seven major plates, including their size, motion, and the types of boundaries between them.
- Describe the geological processes that take place at divergent and convergent plate boundaries, and explain why transform faults exist
- Explain how supercontinents form and how they break apart.
- Explain why tectonic plates move.

Plate tectonics is the model or theory that we use to understand how our planet works: it explains the origins of continents and oceans, the origins of folded rocks and mountain ranges, the presence of different kinds of rocks, the causes and locations of earthquakes and volcanoes, and changes in the positions of continents over time. So... everything!

The theory of plate tectonics was proposed to the geological community more than 100 years ago, so it may come as a surprise that an idea underpinning the study of Earth today did not become an accepted part of geology until the 1960s. It took many decades for this theory to become accepted for two main reasons. First, it was a radically different perspective on how Earth worked, and geologists were reluctant to entertain an idea that seemed preposterous in the context of the science of the day. The evidence and understanding of Earth that would have supported plate tectonic theory simply didn't exist until the mid-twentieth century. Second, their opinion was affected by their view of the main proponent, Alfred Wegener. Wegener was not trained as a geologist, so he lacked credibility in the eyes of the geological community. Alfred Wegener was also German, whereas the geological establishment was centred in Britain and the United States- and Britain and the United States were at war with Germany in the first part of the 20th century. In summary, plate tectonics was an idea too far ahead of its time, and delivered by the wrong messenger.

4.1 Alfred Wegener's Arguments for Plate Tectonics

Alfred Wegener (1880-1930; Figure 4.2) earned a PhD in astronomy at the University of Berlin in 1904, but had a keen interest in geophysics and meteorology, and focused on meteorology for much of his academic career.

Figure 4.2 Alfred Wegener during a 1912-1913 expedition to Greenland. *Source: Alfred Wegener Institute (2008) Public Domain*

In 1911 Wegener happened upon a scientific publication that described matching Permian-aged terrestrial fossils in various parts of South America, Africa, India, Antarctica, and Australia. He concluded that because these organisms could not have crossed the oceans to get from one continent to the next, the continents must have been joined in the past, permitting the animals to move from one to the other (Figure 4.3). Wegener envisioned a supercontinent made up of all the present day continents, and named it **Pangea** (meaning "all land"). He described the motion of the continents reconfiguring themselves as **continental drift**.

Figure 4.3 The distribution of several Permian terrestrial fossils that are present in various parts of continents now separated by oceans. During the Permian, the supercontinent Pangea included the supercontinent Gondwana, shown here, along with North America and Eurasia. *Source: J.M. Watson, USGS (1999) Public Domain.*

Wegener pursued his idea with determination, combing libraries, consulting with colleagues, and making observations in an effort to find evidence in support of it. He relied heavily on matching geological patterns across oceans, such as sedimentary strata in South America matching those in Africa, North American coalfields matching those in Europe, the mountains of Atlantic Canada matching those of northern Britain—both in structure and rock type—and comparisons of rocks in the Canadian Arctic with those of Greenland (Figure 4.4).

Wegener also called upon evidence for the Carboniferous and Permian (~300 Ma) Karoo Glaciation from South America, Africa, India, Antarctica, and Australia (Figure 4.5). He argued that this could only have happened if these continents were once all connected as a single supercontinent. He also cited evidence (based on his own astronomical observations) that showed that the continents were moving with respect to each other, and determined a separation rate between Greenland and Scandinavia of 11 m per year, although he admitted that the measurements were not accurate. (The separation rate is actually about 2.5 cm per year.)

Wegener first published his ideas in 1912 in a short book called *Die Entstehung der Kontinente* (*The Origin of Continents*), and then in 1915 in *Die Entstehung der Kontinente und Ozeane* (*The Origin of Continents and Oceans*). He revised this book several times up to 1929. It was translated into French, English, Spanish, and Russian in 1924.

The main criticism of Wegener's idea was that he could not explain how continents could move. Remember that, as far as anyone was concerned, Earth's crust was continuous, not broken into plates. Thus, any mechanism Wegener could think of would have to fit with that model of Earth's structure. Geologists at the time were aware that continents were made of different rocks than the ocean crust, and that the material making up the continents was less dense, so Wegener proposed that the continents were like icebergs floating on the heavier ocean crust. He suggested that the continents were moved by the effect of Earth's rotation pushing objects toward the equator, and by the lunar and solar tidal forces, which tend to push objects toward the west. However, it was quickly shown that these forces were far too weak to move continents, and without any reasonable mechanism to make it work, Wegener's theory was quickly dismissed by most geologists of the day.

Figure 4.4 Diagram from Alfred Wegener's book *Die Entstehung der Kontinente und Ozeane* comparing rock types on Canadian Arctic Islands and Greenland. *Source: Karla Panchuk (2018) CC BY 4.0. See Appendix C for more attributions.*

Figure 4.5 Carboniferous and Permian Karoo Glaciation in the southern hemisphere. Paleogeographic reconstruction for 306 million years ago. *Source: Cropped from C. R. Scotese, PALEOMAP Project (www.scotese.com) view source. See Appendix C for terms of use.*

Alfred Wegener died in Greenland in 1930 while carrying out studies related to glaciation and climate. At the time of his death, his ideas were tentatively accepted by a small minority of geologists, and firmly rejected by most. But within a few decades that was all to change.

4.2 Global Geological Models of the Early 20th Century

The untimely death of Alfred Wegener did not solve any problems for those who opposed his ideas, because they still had some inconvenient geological truths to deal with. One of those was explaining the distribution of terrestrial species across five continents that are currently separated by hundreds or thousands of kilometres of ocean water, and another was explaining the origin of extensive fold-belt mountains, such as the Appalachians, the Alps, the Himalayas, and the Canadian Rockies.

Before we continue, it is important to know what was generally believed about global geology before plate tectonics. At the beginning of the 20th century, geologists had a good understanding of how most rocks were formed and understood their relative ages through interpretation of fossils, but there was considerable controversy regarding the origin of mountain chains, especially fold-belt mountains. At the end of the 19th century, one of the prevailing views on the origin of mountains was the theory of **contractionism** — the idea that since Earth is slowly cooling, it must also be shrinking. In this scenario, mountain ranges had formed like the wrinkles on a dried-up apple. Oceans formed above parts of former continents that had settled downward and become submerged.

While this hypothesis helped to address the dilemma of the terrestrial fossils by explaining how continents once connected could now be separated by oceans, it came with its own set of problems. One problem was that Earth wasn't cooling fast enough to create the necessary amount of shrinking. Another problem was the principle of isostasy (already understood for several decades; see Section 3.5 for a review of isostasy), which wouldn't allow blocks of continental crust to sink in the way necessary for oceans to form in accordance with contractionist theory.

Another widely held view was **permanentism**, the idea that the continents and oceans have always been generally the same as they are today. This view incorporated a mechanism for creation of mountain chains known as the **geosyncline** theory. A geosyncline is a thick (potentially 1000s of metres) deposit of sediments and sedimentary rocks, typically situated along the edge of a continent, and derived from continental weathering (Figure 4.6).

Figure 4.6 The development of a geosyncline along a continental margin. (Note that a geosyncline is not related to a syncline, which is a downward fold in rocks.) *Source: Steven Earle (2015) CC BY 4.0*

The idea that geosynclines developed into fold-belt mountains originated in the middle of the 19th century. It was first proposed by James Hall and later elaborated upon by Dwight Dana, both of whom worked extensively in the Appalachian Mountains of the eastern United States. The process of converting a geosyncline into a mountain belt was believed to involve compression by forces pushing from either side, causing sedimentary layers within the geosyncline to fold up. In 1937, Philip Kuenen published a paper of experiments with layers of paraffin wax to test how this might work. He was able to cause layers within a geosyncline to fold up as the geosyncline deepened and became more tightly folded during the experiment (Figure 4.7).

Figure 4.7 Simulation of mountain building within a geosyncline using layers of wax. Left- A sequence of photographs showing deformation in the wax layers as pistons apply increasing amounts of compression from the side. Right- Close-up view of slices through the wax layers at the end of the experiment, showing that stiffer white layers of wax folded in a way that resembled the folds in mountain belts. *Source: Karla Panchuk (2018) CC BY 4.0. Photographs from Kuenen (1937) Public Domain*

The problem with the geosynclinal hypothesis for mountain building is that the lateral forces required to cause the compression were never adequately explained. Kuenen compressed the wax layers in his experiment by using pistons that pushed horizontally from either side, but described that mechanism as unrealistic. He explained that the pistons were just to get the process started within the experiment, and that in nature the main force was likely that of gravity pulling the geosyncline downward, drawing the sides together as it folded. When the sides were drawn together, this provided the compression to fold the sediments within the geosyncline. He couldn't specify how the process got started in nature, but suggested that there could be a variety of reasons for an irregularity in the crust to respond to forces in such a way that would trigger downward sagging and folding.

Proponents of the geosyncline theory of mountain formation—and there were many well into the 1960s—also had the problem of explaining the intercontinental terrestrial fossil matchups. The explanation offered was that land bridges had once linked the continents, permitting animals and plants to migrate back and forth. One proponent of this idea was the American naturalist Ernest Ingersoll. Referring to evidence of past climate changes, Ingersoll contributed the following to the 1918-20 edition of the *Encyclopedia Americana*:

> *The most interesting feature of these changes, however, is that by which, now and again, the Old World was connected with the New by necks or spaces of land, known as "land-bridges"; especially as these permitted an interchange of plants and animals, giving to us many new ones from the other side of the ocean, including, finally, man himself.*

A problem with the land-bridge hypothesis is that there is no evidence of land bridges that could account for the fossil distribution patterns. The world's oceans are approximately 4 km deep on average, so the underwater slopes leading up to a land bridge would have to have been at least 10s of km wide in most places, and many times that in others. Even if flooded, a land bridge of that size would still be visible in the shape of ocean-floor terrain. Isostasy would not permit such a land bridge to sink down without leaving a trace.

We do know the locations of some past land bridges, but they were very different from the ones that would be required for this hypothesis. They are bridges such as the flooded Bering Strait land bridge, which is beneath only 30 to 50 m of water, and was exposed when sea levels were much lower because of water being locked in polar ice caps during the last major glaciation event. The narrowest point of the Bering Strait is 82 km wide. The shortest distance between South America and Africa is more than 2800 km.

Exercise: Fitting the Continents Together

The main continents around the Atlantic Ocean are shown in Figure 4.8 with the shapes that they might have had during the Mesozoic, including the extents of their continental shelves. Cut these shapes out and see how well you can fit them together in the positions that these areas occupied within Pangea. Refer to a map of Pangea to help you make the fit. Note that the fit of the continents is even better than this, as distortions are introduced when rendering Earth onto a flat map. A better fit could be accomplished if you were to do this exercise upon the surface of a globe.

Figure 4.8 Mesozoic continent shapes. *Source: Steven Earle (2015) CC BY 4.0*

4.3 Geological Renaissance of the Mid-20th Century

Two key areas of research ultimately led to the acceptance of continental drift, and the formulation of plate tectonic theory. One was the study of **paleomagnetism**, the record of Earth's magnetic field through time. The other was exploration of the ocean floor.

Paleomagnetism (Remnant Magnetism)

When rocks form, some of the minerals that make them up can become aligned with the Earth's magnetic field, just like a compass needle pointing to north. This happens to the mineral magnetite (Fe_3O_4) when it crystallizes from magma. Once the rock cools the crystals are locked in place. This means that if the rock moves, the crystals can't realign themselves, and they retain a **remnant magnetism**. This would be like jamming your compass needle so that if you turned away from north, the needle would turn with you rather than continuing to point north.

Rocks like basalt, which cool from a high temperature and commonly have relatively high levels of magnetite, are particularly susceptible to being magnetized in this way. However, even sediments and sedimentary rocks can take on remnant magnetism as long as they have small amounts of magnetic minerals, because the magnetic grains can gradually become lined up with Earth's magnetic field as the sediments are deposited.

By studying both the horizontal and vertical components of the remnant magnetism, one can tell not only the direction to magnetic north at the time of the rock's formation, but also the latitude where the rock formed relative to magnetic north. Remember that the vertical component of the magnetic field points more sharply downward the closer it is to the magnetic north pole. Figure 4.9 shows the vertical component of remnant magnetism in a sequence of rocks. Notice that the arrow starts out at 500 Ma pointing slightly upward. This means that the rocks were in the southern hemisphere. As the rocks get younger, the arrow tilts toward horizontal, and then points downward. This indicates that the rocks were getting progressively closer to the north magnetic pole.

Figure 4.9 Rock layers recording remnant magnetism. The red arrows represent the direction of the vertical component of Earth's magnetic field. The oldest rock has a magnetic dip characteristic of the southern hemisphere, but over time the dip changes, indicating that the rocks moved toward magnetic north. *Source: Steven Earle (2015) CC BY 4.0*

Apparent Polar Wandering Paths

In the early 1950s, a group of geologists from Cambridge University, including Keith Runcorn, Ted Irving,[1] and several others, started looking at the remnant magnetism of Phanerozoic British and European volcanic rocks, and collecting **paleomagnetic** data. Using an analysis similar to that in Figure 4.9, they noticed that rocks of different ages sampled from the same general area showed very different magnetic pole positions (the green line in Figure 4.10). They assumed this meant that Earth's magnetic pole had moved around significantly over time along **polar wandering paths**, rather than staying close to the geographic north pole as it does today. At the time, geophysical models suggested that the magnetic poles did not need to be aligned with the rotational poles, so this wasn't an unreasonable conclusion, given what was known.

Figure 4.10 Apparent polar-wandering paths (APWP) for Eurasia and North America. The view is from the geographic north pole (black dot) looking down. Dots along each path show the location of magnetic north as determined from paleomagnetic data. Left- Data from Eurasia and North America agree on the location of magnetic north today (time 0), but not at any time in the past. Right- Once continent motion has been accounted for, there is agreement in data from Eurasia and North America on the location of magnetic north over the past 500 million years. *Source: Steven Earle (2015) CC BY 4.0*

Runcorn and colleagues extended their paleomagnetic studies to North America, and began to realize that their initial conclusion had a problem. Notice that on the right of Figure 4.10 the polar wandering path for North America (in red) does not match the path for Eurasia (in green). For example, data from North America suggest that 200 Ma ago, magnetic north was somewhere in China, whereas data from Europe said it was in the Pacific Ocean. There could only have been one magnetic north pole position at 200 Ma, therefore the only way to explain the discrepancy was if Europe and North America moved along different paths during this time while the pole stayed in more or less the same location.

---

[1] Ted Irving later set up a paleomagnetic lab at the Geological Survey of Canada in Sidney BC, and did important work on understanding the geology of western North America

The polar wandering paths were not actually records of the pole moving, they just looked that way, so the paths are now referred to as *apparent* polar wandering paths (APWP). Subsequent paleomagnetic work showed that unique apparent polar wandering paths can be derived from rocks in South America, Africa, India, and Australia. In 1956, Runcorn became a proponent of continental drift. There was simply no other way to explain the data.

This paleomagnetic work of the 1950s was the first new evidence in favour of continental drift, and it led a number of geologists to start thinking that the idea might have some merit. Nevertheless, for a majority of geologists, this type of evidence was not sufficiently convincing to get them to change their views.

Ocean Basin Geology and Geography

During the 20th century, our knowledge and understanding of the ocean basins and their geology increased dramatically. Before 1900 we knew virtually nothing about the bathymetry (the hills and valleys of the ocean floor) and geology of the oceans. By the end of the 1960s, we had detailed maps of the topography of the ocean floors, a clear picture of the geology of ocean floor sediments and the solid rocks beneath them, and almost as much information about the geophysical nature of ocean rocks as of continental rocks.

Acoustic Depth Sounding

Up until the 1920s, ocean depths were measured using weighted lines dropped overboard. In deep water this is a painfully slow process and the number of soundings in the deep oceans was probably fewer than 1,000. That is roughly one depth sounding for every 350,000 square kilometres of the ocean. To put that in perspective, it would be like trying to describe the topography of British Columbia with elevation data for only a half a dozen points!

Figure 4.11 A ship-borne acoustic depth sounder. The instrument emits sound (black arcs) that reflects off the sea floor and returns to the surface (white arcs). The time interval between emitting the sound and detecting it on receivers on the ship is proportional to the water depth. *Source: Steven Earle (2015) CC BY 4.0*

The voyage of the *Challenger* in 1872 and the laying of trans-Atlantic cables had shown that there were mountains beneath the seas, but most geologists and oceanographers still believed that the oceans were essentially vast basins with flat bottoms, filled with thousands of metres of sediments.

Following development of acoustic depth sounders (Figure 4.11) in the 1920s, the number of depth readings increased by many orders of magnitude, and by the 1930s there was no doubt that major mountain chains ran through all of the world's oceans. During and after World War II, there was a well-organized campaign to study the oceans, and by 1959, sufficient bathymetric data had been collected to produce detailed maps of all the oceans (Figure 4.12).

Figure 4.12 Ocean floor bathymetry (and continental topography). Inset (a): the mid-Atlantic ridge, (b): the Newfoundland continental shelf, (c): the Nazca trench adjacent to South America, and (d): the Hawaiian Island chain. *Source: Steven Earle (2015) CC BY 4.0; Base map after NOAA (2006) Public Domain*

The important physical features of the ocean floor are:

- Extensive linear ridges (commonly in the central parts of the oceans) at depths of 2,000 to 3,000 m (Figure 4.12, a)
- Fracture zones perpendicular to the ridges (Figure 4.12, a)
- Deep-ocean plains at depths of 4,000 to 5,000 m (Figure 4.12, a and d)
- Relatively flat and shallow continental shelves with depths under 500 m (Figure 4.12, b)
- Deep trenches (up to 11,000 m deep), most near the continents (Figure 4.12, c)
- Seamounts and chains of seamounts (Figure 4.12, d)

Seismic Reflection Sounding

Seismic reflection sounding involves transmitting high-energy sound bursts and then measuring the echoes with a series of receivers called **geophones** towed behind a ship. The technique is related to acoustic sounding as described above, however, much more energy is transmitted and the sophistication of the data processing is much greater. As the technique evolved, and the amount of energy used was increased, it became possible to see *through* the sea-floor sediments and map the bedrock topography and crustal thickness. This allowed sediment thicknesses to be mapped (Figure 4.13).

Figure 4.13 Map of global sediment thickness. *Source: NOAA (2003) Public Domain*

It was soon discovered that although the sediments were up to several 1000s of m thick near the continents, they were relatively thin — or even non-existent — along ocean ridges (Figure 4.14). The seismic studies also showed that the crust is relatively thin under the oceans (5 km to 6 km) compared to the continents (30 km to 60 km) and geologically very consistent, composed almost entirely of basalt.

Figure 4.14 Topographic section at an ocean ridge based on reflection seismic data. Sediments are not thick enough to be detectable near the ridge, but get thicker on either side. The diagram represents approximately 50 km width, and has a 10x vertical exaggeration. *Source: Steven Earle (2015) CC BY 4.0*

Heat Flow Rates

In the early 1950s, Edward Bullard—who spent time at the University of Toronto but is mostly associated with Cambridge University—developed a probe for measuring the flow of heat from the ocean floor. Bullard and colleagues found higher than average heat-flow rates along the ridges, and lower than average rates in trenches. These data were interpreted as evidence of mantle convection, with areas of high heat flow corresponding to upward convection of hot mantle material, and areas of low heat flow corresponding to downward convection.

Earthquake Belts

With developments of networks of seismographic stations in the 1950s, it became possible to plot the locations *and* depths of both major and minor earthquakes with great accuracy. A remarkable correspondence was observed between earthquake locations and both the mid-ocean ridges and the deep ocean trenches. In 1954 Gutenberg and Richter showed that the ocean-ridge earthquakes were all relatively shallow, and confirmed what had first been shown by Benioff in the 1930s, that earthquakes in the vicinity of ocean trenches were both shallow and deep, but that the deeper ones were situated progressively farther inland from the trenches (Figure 4.15).

Figure 4.15 Aleutian Island subduction zone earthquakes. Left- Map view with earthquakes marked as dots. Red dots are the shallowest earthquakes and blue are the deepest. Quakes get deeper further inland from the trench. Right- Cross-section through a-b. Coloured dots show the depth of earthquakes. Colours correspond to dots in the left figure. Earthquake depth is related to the position of the Pacific plate as it travels beneath the North American plate. *Source: Steven Earle (2015) CC BY 4.0*

Magnetic Stripes on the Sea Floor

In the 1950s, scientists from the Scripps Oceanographic Institute in California persuaded the United States Coast Guard to include **magnetometer** readings on one of their expeditions to study ocean floor topography. The first comprehensive magnetic data set was compiled in 1958 for an area off the coasts of BC and Washington State. This survey revealed a bewildering pattern of low and high magnetic field intensity in sea-floor rocks (Figure 4.16). When the data were first plotted on a map in 1961, nobody understood them — not even the scientists who collected them. Many 1000s of km of magnetic surveys were conducted over the next several years.

The wealth of new data from the oceans began to significantly influence geological thinking in the 1960s. In 1960, Harold Hess from Princeton University advanced a hypothesis with many of the elements that we now accept as **plate tectonics**. He maintained some uncertainty about his proposal however, and in order to deflect criticism from mainstream geologists, he labelled it "geopoetry." In fact, until 1962, Hess didn't even put his ideas in writing — except internally to the U.S. Navy (which funded his research) — but presented them mostly in lectures and seminars.

Figure 4.16 Pattern of sea-floor magnetic field intensity off the west coast of British Columbia and Washington. Black regions have higher than average magnetic field instensity, and white regions have lower than average intensity. *Source: Steven Earle (2015) CC BY-SA 4.0, modified after U. S. Geological Survey (n.d.) Public Domain, (adapted from Raff and Mason, 1961).*

Hess proposed that new sea floor was generated from mantle material at the ocean ridges, and that old sea floor was dragged down at the ocean trenches and re-incorporated into the mantle. He suggested that the process was driven by mantle convection currents, rising at the ridges and descending at the trenches (Figure 4.17). He also suggested that the less-dense continental crust did not descend with oceanic crust into trenches, but that colliding landmasses were thrust up to form mountains.

Hess's hypotheses formed the basis for our ideas on **sea-floor spreading** and **continental drift,** but did not go so far as to claim that the crust is made up of separate plates. The Hess model was not roundly criticized, but also not widely accepted, partly because evidence was still lacking.

Figure 4.17 A representation of Harold Hess's model for sea-floor spreading and subduction.
Source: Steven Earle (2015) CC BY 4.0

Collection of magnetic data from the oceans continued in the early 1960s, but the striped patterns remained unexplained. Some assumed that, as with continental crust, the stripes were related to compositional variations in rock, such as variations in the amount of magnetite. The first real understanding of the significance of the striped anomalies was the interpretation by Fred Vine, a Cambridge graduate student. Vine was examining magnetic data from the Indian Ocean and, like others before him, noted that the magnetic patterns were symmetrical on either side of the ridge.

At the same time, other researchers led by groups in California and New Zealand were studying the phenomenon of reversals in Earth's magnetic field. They were trying to determine when such reversals had taken place over the past several million years by analyzing the magnetic characteristics of hundreds of samples from basaltic flows. As discussed in Chapter 3, Earth's magnetic field periodically weakens, then becomes virtually non-existent before becoming re-established with the reverse polarity. During periods of reversed polarity, a compass would point south instead of north.

The time scale of magnetic reversals is irregular. The present "normal" event, known as the Bruhnes magnetic chron, has persisted for about 780,000 years. It was preceded by a 190,000-year reversed event; a 50,000-year normal event known as Jaramillo; and then a 700,000-year reversed event (see Figure 3.16).

In a paper published in September 1963, Vine and his PhD supervisor Drummond Matthews proposed that the patterns associated with ridges were related to the magnetic reversals, and that oceanic crust created from cooling basalt during a *normal* event would have polarity aligned with the present magnetic field, and would produce a positive anomaly (a black stripe on the sea-floor magnetic map). Oceanic crust created during a *reversed* event would have polarity opposite Earth's present field and thus produce a negative magnetic anomaly (a white stripe). The same idea had been put forward a few months earlier by Lawrence Morley, of the Geological Survey of Canada. However, Morley's papers submitted earlier in 1963 to *Nature* and *The Journal of Geophysical Research* were rejected. The idea is sometimes referred to as the Vine-Matthews-Morley (VMM) hypothesis.

Vine, Matthews, and Morley were the first to show this type of correspondence between the relative widths of the stripes and the durations of the magnetic reversals. The VMM hypothesis was confirmed within a few years when magnetic data were compiled from spreading ridges around the world. It was shown that the same general magnetic patterns were present straddling each ridge, although the widths of the anomalies varied according to the spreading rates characteristic of the different ridges. It was also shown that the patterns corresponded with the known timeline of Earth's magnetic field reversals.

In 1963, J. Tuzo Wilson of the University of Toronto proposed the idea of a **mantle plume** or **hot spot** — a place where hot mantle material rises in a stationary and semi-permanent plume, and affects the overlying crust. He based this hypothesis partly on the distribution of the Hawai'ian and Emperor Seamount island chains in the Pacific Ocean (Figure 4.18). The volcanic rock making up these islands becomes progressively younger toward the southeast, culminating in the island of Hawai'i itself, which consists of rock that is almost all younger than 1 Ma.

Figure 4.18 The ages of the Hawai'ian Islands and the Emperor Seamounts in relation to the location of the Hawai'ian mantle plume. *Source: Steven Earle (2015) CC BY 4.0; Base map from the National Geophysical Data Centre/USGS (2005) Public Domain*

Wilson suggested that a stationary plume of hot upwelling mantle material is the source of the Hawaiian volcanism, and that the ocean crust of the Pacific Plate is moving toward the northwest over this hot spot. Near the Midway Islands, the chain makes a pronounced change in direction, from northwest-southeast for the Hawai'ian Islands, to nearly north-south for the Emperor Seamounts. This change has been ascribed to a change in direction of the Pacific Plate moving over the stationary mantle plume. An alternative hypothesis is that rather than the Pacific Plate having undergone a sudden change in motion, the plume itself has moved at least 2,000 km south over the period between 81 and 45 Ma (Tarduno et al., 2003).

There is evidence of many such mantle plumes around the world (Figure 4.19). Most are within ocean basins, including places like Hawai'i, Iceland, and the Galapagos Islands, but some are under continents. One example is the Yellowstone hot spot in the west-central United States, and another is the one responsible for the Anahim Volcanic Belt in central British Columbia. It is evident that mantle plumes are very long-lived phenomena, lasting for at least tens of millions of years, and possibly for hundreds of millions of years in some cases.

Figure 4.19 Mantle plume locations. Selected Mantle plumes: 1: Azores, 3: Bowie, 5: Cobb, 8: Eifel, 10: Galapagos, 12: Hawai'i, 14: Iceland, 17: Cameroon, 18: Canary, 19: Cape Verde, 35: Samoa, 38: Tahiti, 42: Tristan, 44: Yellowstone, 45: Anahim. *Source: Ingo Wölbern (2007) Public Domain*

Oceanic spreading ridges appear to be curved features on Earth's surface, but the ridges are in fact composed of a series of straight-line segments, offset at intervals by faults perpendicular to the ridge (Figure 4.20). In a paper published in 1965, Tuzo Wilson termed these features **transform faults**.

Figure 4.20 Part of the Mid-Atlantic ridge near the equator. Transform faults (red lines) are in between the ridge segments (double white lines), where the yellow arrows (indicating relative plate movement) point in opposite directions. Solid white lines are fracture zones. *Source: Steven Earle (2015) CC BY 4.0*

In the same 1965 paper, Wilson introduced the idea that the crust can be divided into a series of rigid plates, and is thus responsible for the term **plate tectonics**.

Exercise: Paper Transform Fault Model

J. Tuzo Wilson used a paper model similar to the one in Figure 4.21 to explain transform faults to his colleagues. To use this model, print Figure 4.21, cut around the outside, and then slice along the line A-B (the fracture zone) with a sharp knife. Fold down the top half where shown, and then pinch together in the middle. Do the same with the bottom half. When you're done, you should have two folds of paper extending downward as in Figure 4.22.

Above- **Figure 4.21** Transform fault model. Source: Steven Earle (2015) CC BY 4.0, modified after Stewart (1990).

Left- **Figure 4.22** Use of the transform fault model. *Source: Steven Earle (2015) CC BY 4.0*

Exercise: Paper Transform Fault Model, continued

Find someone else to pinch those folds with two fingers just below each ridge crest, and then gently pull apart where shown. As you do, the oceanic crust will emerge from the middle, and you will see that the parts of the fracture zone between the ridge crests will be moving in opposite directions (this is the transform fault) while the parts of the fracture zone outside of the ridge crests will be moving in the same direction. You will also see that the oceanic crust is being magnetized as it forms at the ridge. The magnetic patterns represent the last 2.5 Ma of geological time.

There are other versions of this model available at https://web.viu.ca/earle/transform-model/. For more information see Earle (2004).

4.4 Plates, Plate Motions, and Plate-Boundary Processes

The ideas of continental drift and sea-floor spreading became widely accepted by 1965, and more geologists started thinking in these terms. By the end of 1967, Earth's surface had been mapped into a series of plates (Figure 4.23). The major plates are Eurasian, Pacific, Indian, Australian, North American, South American, African, and Antarctic plates. There are also numerous small plates (e.g., Juan de Fuca, Nazca, Scotia, Philippine, Caribbean), and many very small plates or sub-plates. The Juan de Fuca Plate is actually three separate plates (Gorda, Juan de Fuca, and Explorer), all moving in the same general direction but at slightly different rates.

Figure 4.23 A detailed map of Earth's tectonic plates. *Source: Paul Lowman and Jacob Yates, NASA Goddard Space Flight Center (2002) Public Domain. To see an enlarged version go to* https://visibleearth.nasa.gov/view.php?id=88415

Plate motions can be tracked using Global Positioning System (GPS) data from different locations on Earth's surface. Rates of motions of the major plates range from less than 1 cm/y to more than 10 cm/y. The Pacific Plate is the fastest, moving at more than 10 cm/y in some areas, followed by the Australian and Nazca Plates. The North American Plate is one of the slowest, averaging ~1 cm/y in the south up to almost 4 cm/y in the north.

Plates move as rigid bodies, so it may seem surprising that the North American Plate can be moving at different rates in different places. The explanation is that plates rotate as they move; the North American Plate, for example, rotates counter-clockwise, while the Eurasian Plate rotates clockwise.

Boundaries between the plates are of three types: **divergent** (moving apart), **convergent** (moving together), and **transform** (moving side by side). The plates are made up of crust and lithospheric mantle (Figure 4.24). Even though the plates are in constant motion, and move in different directions, there is never a significant amount of space between them. Plates move along the lithosphere-asthenosphere boundary, because the asthenosphere is relatively weak. It deforms as the plates move, rather than locking them in place.

Figure 4.24 The crust and upper mantle. Tectonic plates consist of lithosphere, which includes the crust and the lithospheric (rigid) part of the mantle. *Source: Steven Earle (2015) CC BY 4.0*

At spreading centres, the lithospheric mantle is relatively thin. The upward convective motion of hot mantle material generates temperatures that are too high for the existence of a significant thickness of rigid lithosphere at the same time that the plates are falling away from each other (Figure 4.17).

The fact that plates include both crustal material and lithospheric mantle material makes it possible for a single plate to be include both oceanic and continental crust. The North American Plate includes most of North America, plus half of the northern Atlantic Ocean. Similarly the South American Plate extends across the western part of the southern Atlantic Ocean, while the European and African plates each include part of the eastern Atlantic Ocean. The Pacific Plate is almost entirely oceanic, but it does include the part of California west of the San Andreas Fault.

Divergent Boundaries

Divergent boundaries are spreading boundaries, where new oceanic crust is created from magma derived from partial melting of the mantle. The partial melting happens when hot mantle rock is moved from deep within Earth where pressures are too high for it to be liquid, to shallower depths where the pressure is much lower (Figure 4.25, bottom left).

The triangular zone of partial melting near the ridge crest is approximately 60 km thick and the proportion of magma is about 10% of the rock volume, thus producing crust that is about 6 km thick once the melt escapes from the rock in which it formed, and ascends. Most divergent boundaries are located in the oceans, and the crustal material created at a spreading boundary is always oceanic in character; in other words, it is mafic igneous rock (basalt or gabbro, with minerals rich in iron and magnesium). Spreading rates vary considerably, from 1 cm/y to 3 cm/y in the Atlantic, to between 6 cm/y and 10 cm/y in the Pacific. Some of the processes taking place in this setting include (Figure 4.25, top):

- Melted rock (magma) from the mantle rising up to fill the voids left by divergence of the two plates
- **Pillow lavas** forming where melted rock emerges on the ocean floor and is cooled by seawater (Figure 4.25, bottom right)
- Vertical sheeted dykes intruding into cracks resulting from the spreading
- Magma cooling more slowly in the lower part of the new crust, forming bodies of gabbro

Figure 4.25 Divergent boundary. Lower left- General processes taking place along divergent boundaries. Top- Expanded view of the white box showing divergent boundary processes and materials. Bottom right- Pillow basalts from the ocean floor of Hawai'i. *Source: Lower left- Steven Earle (2015) CC BY 4.0; Top- Steven Earle (2015) CC BY 4.0, modified after Sinton and Detrick (1992); Lower right- NOAA (1988) Public Domain*

Spreading is thought to start with lithosphere being warped upward into a dome by buoyant material from an underlying mantle plume or series of mantle plumes. The buoyancy of the mantle plume causes the dome to fracture in a radial pattern, with three arms spaced at approximately 120° (Figure 4.26).

Figure 4.26 Depiction of the process of dome and three-part rift formation (left) and of continental rifting between the African and South American parts of Pangea at around 200 Ma (right) *Source: Steven Earle (2015) CC BY 4.0*

When a series of mantle plumes exists beneath a large continent, the resulting rifts may align and lead to the formation of a rift valley, such as the present-day Great Rift Valley in eastern Africa. This type of valley may eventually develop into a linear sea (such as the present-day Red Sea), and finally into an ocean (such as the Atlantic). It is likely that as many as 20 mantle plumes, many of which still exist, were responsible for the initiation of the rifting of Pangea along what is now the mid-Atlantic ridge (see the Atlantic Ocean mantle plume locations in Figure 4.19).

Convergent Boundaries

Convergent boundaries, where two plates are moving toward each other, are of three types, depending on whether ocean or continental crust is present on either side of the boundary. The types are ocean-ocean, ocean-continent, and continent-continent.

Ocean-Ocean Convergent Boundaries

At an ocean-ocean convergent boundary, a plate margin consisting of oceanic crust and lithospheric mantle is **subducted**, or travels beneath, the margin of the plate with which it is colliding (Figure 4.27). Often it is the older and colder plate that is denser and subducts beneath the younger and hotter plate. Ocean trenches commonly form along these boundaries.

Figure 4.27 Configuration and processes of an ocean-ocean convergent boundary
Source: Steven Earle (2015) CC BY 4.0

As the subducting crust is heated and the pressure increases, water is released from within the subducting material. This water comes primarily from alteration of the minerals pyroxene and olivine to serpentine near the spreading ridge shortly after the rock's formation. The water mixes with the overlying mantle, which lowers the melting point of mantle rocks, causing magma to form. This process is called **flux melting** or **fluid-induced melting**.

The newly produced magma, which is lighter than the surrounding mantle rocks, rises through the mantle and sometimes through the overlying oceanic crust to the ocean floor where it creates a chain of volcanic islands known as an island arc. A mature island arc develops into a chain of relatively large islands (such as Japan or Indonesia) as more and more volcanic material is extruded and sedimentary rocks accumulate around the islands. The largest earthquakes occur near the surface where the subducting plate is still cold and strong.

Examples of ocean-ocean convergent zones are subduction of the Pacific Plate south of Alaska (Aleutian Islands) and west of the Philippines, subduction of the Indian Plate south of Indonesia, and subduction of the Atlantic Plate beneath the Caribbean Plate.

Ocean-Continent Convergent Boundaries

At an ocean-continent convergent boundary, the oceanic plate is subducted beneath the continental plate in the same manner as at an ocean-ocean boundary. Rocks and sediment on the continental slope are thrust up into an **accretionary wedge**, and compression leads to faults forming within the continental plate (Figure 4.28). The mafic magma produced adjacent to the subduction zone rises to the base of the continental crust and leads to partial melting of the crustal rock. The resulting magma ascends through the crust, producing a mountain chain with many volcanoes.

Figure 4.28
Configuration and processes of an ocean-continent convergent boundary Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0

Examples of ocean-continent convergent boundaries are subduction of the Nazca Plate under South America (which has created the Andes Range) and subduction of the Juan de Fuca Plate under North America (creating the mountains Garibaldi, Baker, St. Helens, Rainier, Hood, and Shasta, collectively known as the Cascade Range).

Continent-Continent Convergent Boundary

A continent-continent collision occurs when a continent or large island that has been moved along with subducting oceanic crust collides with another continent (Figure 4.29). The colliding continental material will not be subducted because it is not dense enough, but the root of the oceanic plate will eventually break off and sink into the mantle. There is tremendous deformation of the pre-existing continental rocks, and creation of mountains from that rock, as well as from any sediments that had accumulated along the shores of both continental masses, and commonly also from some ocean crust and upper mantle material.

Figure 4.29
Configuration and processes of a continent-continent convergent boundary *Source: Steven Earle (2015) CC BY 4.0*

Examples of continent-continent convergent boundaries are the collision of the India Plate with the Eurasian Plate, creating the Himalaya Mountains, and the collision of the African Plate with the Eurasian Plate, creating the series of ranges extending from the Alps in Europe to the Zagros Mountains in Iran.

When a subduction zone is jammed shut by a continent-continent collision, plate tectonic stresses that are still present can sometimes cause a new subduction zone to develop outboard of the colliding plate.

Transform Boundaries

Transform boundaries exist where one plate slides past another without producing or destroying crust, except in the special case where the transform boundary has bends and jogs. There will be collisions and divergence on a small scale as the jogs crash into the bends, or open up small windows to deeper crust.

Most transform faults connect segments of mid-ocean ridges and are thus ocean-ocean plate boundaries (Figure 4.20). Some transform faults connect continental parts of plates. An example is the San Andreas Fault, which connects the southern end of the Juan de Fuca Ridge with the northern end of the East Pacific Rise (a ridge) in the Gulf of California (Figures 4.30 and 4.31). The part of California west of the San Andreas Fault and all of Baja California are on the Pacific Plate. But transform faults do not just connect divergent boundaries; the Queen Charlotte Fault connects the north end of the Juan de Fuca Ridge, starting at the north end of Vancouver Island, to the Aleutian subduction zone.

Left- **Figure 4.30** The San Andreas Fault extends from the north end of the East Pacific Rise in the Gulf of California to the southern end of the Juan de Fuca Ridge. All of the red lines on this map are transform faults. *Source: Steven Earle (2015) CC BY 4.0*

Below- **Figure 4.31** The San Andreas Fault at Parkfield in central California. The person with the orange shirt is standing on the Pacific Plate and the person at the far side of the bridge is on the North American Plate. The bridge is designed to slide on its foundation. *Source: Steven Earle (2015) CC BY 4.0*

Exercise: A Different Type of Transform Fault

Figure 4.32 shows the Juan de Fuca (JDF) and Explorer plates off the coast of Vancouver Island. The JDF Plate is moving toward the North American Plate at ~ 4.5 cm/y. We think that the Explorer Plate is also moving east, but we don't know the rate, and there is evidence that it is slower than the JDF Plate.

The boundary between the two plates is the Nootka Fault, which is the location of frequent small-to-medium earthquakes (up to magnitude ~5), as depicted by the red stars. Explain why the Nootka Fault is a transform fault, and show the *relative* sense of motion along the fault with two small arrows.

Figure 4.32 Juan de Fuca and Explorer plates separated by the Nootka Fault (marked with red stars). *Source: Steven Earle (2015) CC BY 4.0*

Plate Tectonics and Supercontinent Cycles

The present continents were once all part of a supercontinent that Alfred Wegener named **Pangea** (*all land*). More recent studies of continental matchups and the magnetic ages of ocean-floor rocks have enabled us to reconstruct the history of the break-up of Pangea.

Pangea began to rift apart along a line between Africa and Asia and between North America and South America at around 200 Ma (Figure 4.33). During the same period the Atlantic Ocean began to open up between northern Africa and North America, and India broke away from Antarctica. Between 200 and 150 Ma, rifting started between South America and Africa and between North America and Europe, and India moved north toward Asia. By 80 Ma, Africa had separated from South America, and most of Europe had separated from North America. By 50 Ma, Australia had separated from Antarctica, and shortly after that, India collided with Asia.

Within the past few million years, rifting has occurred in the Gulf of Aden and the Red Sea, and also within the Gulf of California. Incipient rifting has begun along the Great Rift Valley of eastern Africa, extending from Ethiopia and Djibouti on the Gulf of Aden (Red Sea) all the way south to Malawi.

Pangea was not the first supercontinent. It was preceded by Pannotia (600 to 540 Ma), Rodinia (1,100 to 750 Ma), and by others before that. In fact, in 1966, Tuzo Wilson proposed that supercontinents are part of an on-going cycle, which we now refer to as a **Wilson cycle**. In a Wilson cycle, continents break up, and fragments drift apart only to collide again and make a new continent.

Figure 4.33 Sequence of paleogeographic reconstructions showing the breakup of Pangea. *Source: Karla Panchuk (2017) CC BY-NC-SA 4.0. Maps from C. R. Scotese, PALEOMAP Project (www.scotese.com). See Appendix C for map sources and terms of use.*

At present we are in the stages of a Wilson cycle where fragments are drifting and changing their configuration. North and South America, Europe, and Africa are moving with their respective portions of the Atlantic Ocean. The eastern margins of North and South America and the western margins of Europe and Africa are called passive margins because there is no subduction taking place along them. Because the oceanic crust formed by spreading along the mid-Atlantic ridge is not currently being subducted (except in the Caribbean), the Atlantic Ocean is slowly getting bigger, and the Pacific Ocean is getting smaller.

This situation may not continue for too much longer, however. As the Atlantic Ocean floor gets weighed down around its margins by great thickness of continental sediments, it will be pushed farther and farther into the mantle, and eventually the oceanic lithosphere may break away from the continental lithosphere and begin to subduct (Figure 4.34).

Figure 4.34 Development of a subduction zone at a passive margin. Times A, B, and C are separated by tens of millions of years. Once the oceanic crust breaks off and starts to subduct, the continental crust (North America in this case) may no longer be pushed to the west and could start to move east because the rate of spreading in the Pacific basin is faster than along the Mid-Atlantic Ridge. *Source: Steven Earle (2015) CC BY 4.0*

A subduction zone will develop, and the oceanic plate will begin to descend under the continent. Once this happens, the continents will no longer continue to move apart because the spreading at the mid-Atlantic ridge will be taken up by subduction. If spreading along the mid-Atlantic ridge continues to be slower than spreading within the Pacific Ocean, the Atlantic Ocean will start to close up, and eventually (in a 100 million years or more) North and South America will collide again with Europe and Africa. If this continues without changing for another few hundred million years, we will be back to where we started, with one supercontinent (Figure 4.35).

Figure 4.35 Sequence of reconstructions showing the possible future configuration of land masses on Earth at 50, 150, and 250 million years from now. Movements culminate in the formation of a new supercontinent called Pangea Ultima. *Source: Karla Panchuk (2017) CC BY-NC-SA 4.0. Maps from C. R. Scotese, PALEOMAP Project (www.scotese.com). See Appendix C for map sources and terms of use.*

There is strong evidence around the margins of the Atlantic Ocean that this process has taken place before. There are roots of ancient mountain belts along the eastern margin of North America, the western margin of Europe, and the north-western margin of Africa, which show that these landmasses once collided with each other to form a mountain chain. The mountain chain might have been as big as the Himalayas.

The apparent line of collision runs between Norway and Sweden, between Scotland and England, through Ireland, through Newfoundland and the Maritimes, through the north-eastern and eastern states, and across the northern end of Florida. When rifting of Pangea started at approximately 200 Ma, the fissuring was along a different line from the line of the earlier collision. This is why some of the mountain chains formed during the earlier collision can be traced from Europe to North America and from Europe to Africa.

It is probably no coincidence that the Atlantic Ocean rift may have occurred in approximately the same place during two separate events several hundred million years apart. The series of hot spots that has been identified in the Atlantic Ocean may also have existed for several hundred million years, and thus may have contributed to rifting in roughly the same place on at least two separate occasions (Figure 4.36).

Figure 4.36 A scenario for the Wilson cycle. The cycle starts with continental rifting above a series of mantle plumes (red dots, A). The continents separate (B), and then re-converge some time later, forming a fold-belt mountain chain. Eventually rifting is repeated, possibly because of the same set of mantle plumes (D), but this time the rift is in a different place. *Source: Steven Earle (2015) CC BY 4.0*

4.5 Mechanisms for Plate Motion

Mantle convection is often said to be critical to plate tectonics. While this is almost certainly so, there is still debate about the actual forces that make the plates move. One side of the argument holds that the plates are only moved by the traction caused by mantle convection, and that friction between the asthenosphere and lithosphere pulls the lithosphere along as the mantle convects. The other side holds that traction plays only a minor role and that **ridge-push** and **slab-pull** are more important (Figure 4.37).

Ridge-push refers to gravity causing lithosphere to slide downhill away from the elevated mid-ocean ridges. **Slab-pull** refers to the weight of subducting slabs dragging the rest of the plate down into the mantle.

Figure 10.35 Figure 4.37 Models for plate motion mechanisms. *Source: Steven Earle (2015) CC BY 4.0*

Kearey and Vine (1996) have listed some compelling arguments in favour of the ridge-push/slab-pull model:

- Plates that are attached to subducting slabs (e.g., Pacific, Australian, and Nazca Plates) move the fastest, and plates that are not (e.g., North American, South American, Eurasian, and African Plates) move significantly slower.

- In order for the traction model to apply, the mantle would have to be moving about five times faster than the plates are moving because the coupling between the partially liquid asthenosphere and the plates is not strong. Such high rates of convection are not supported by geophysical models.

- Although large plates have the potential for much higher convection traction, plate velocity is not related to plate area.

Chapter 4 Summary

The topics covered in this chapter can be summarized as follows:

4.1 Alfred Wegener's Arguments for Plate Tectonics

The evidence for continental drift in the early 20th century included the matching of continental shapes on either side of the Atlantic, and the geological and fossil matchups between continents that are now thousands of kilometres apart.

4.2 Global Geological Models of the Early 20th Century

The established theories of global geology were permanentism and contractionism, but neither of these theories was able to explain some of the evidence that supported the idea of continental drift.

4.3 Geological Renaissance of the Mid-20th Century

Giant strides were made in understanding Earth during the middle decades of the 20th century, including discovering magnetic evidence of continental drift, mapping the topography of the ocean floor, describing the depth relationships of earthquakes along ocean trenches, measuring heat flow differences in various parts of the ocean floor, and mapping magnetic reversals on the sea floor. By the mid-1960s, the fundamentals of the theory of plate tectonics were in place.

4.4 Plates, Plate Motions, and Plate-Boundary Processes

Earth's lithosphere is made up of over 20 plates that are moving in different directions at rates of between 1 cm/y to greater than 10 cm/y. The three types of plate boundaries are divergent (plates moving apart and new crust forming), convergent (plates moving together and one possibly being subducted), and transform (plates moving side by side). Divergent boundaries form where existing plates are rifted apart, and it is hypothesized that this is caused by a series of mantle plumes. Subduction zones can form where accumulation of sediment at a passive margin leads to separation of oceanic and continental lithosphere. Supercontinents form and break up through these processes.

4.5 Mechanisms for Plate Motion

It is widely believed that ridge-push and slab-pull are the main mechanisms for plate motion, as opposed to traction by mantle convection. Mantle convection is a key factor for producing the conditions necessary for ridge-push and slab-pull.

Review Questions

1. List some of the evidence used by Wegener to support his idea of moving continents.
2. What was the primary weakness in Wegener's continental drift hypothesis?
3. How were mountains thought to be formed (a) by contractionists and (b) by permanentists?
4. How were the trans-Atlantic paleontological matchups explained in the late 19th century?
5. How did we learn about the topography of the sea floor in the early part of the 20th century?
6. What evidence from paleomagnetic studies provided support for continental drift?
7. Which parts of the oceans are the deepest?
8. Why is there less sediment along ocean ridges than on other parts of the sea floor?
9. How were the oceanic heat-flow data related to mantle convection?
10. Describe the spatial distribution of earthquakes at ocean ridges and ocean trenches.

11. In the model for ocean basins developed by Harold Hess, what happens at oceanic ridges and what happens at oceanic trenches?
12. What aspect of plate tectonics was not included in the Hess model?
13. What is a mantle plume and what is its expected lifespan?
14. Describe the nature of movement at an ocean ridge transform fault (a) between the ridge segments, and (b) outside of ridge segments.
15. How is it possible for a plate to include both oceanic and continental crust?
16. What is the likely relationship between mantle plumes and the development of a continental rift?
17. Why does subduction not take place at a continent-continent convergence zone?
18. Where are Earth's most recent sites of continental rifting and creation of new ocean floor?
19. What geological situation might eventually lead to the generation of a subduction zone at a passive ocean-continent boundary such as the eastern coast of North America?

Answers at the end of the chapter

References

Earle, S. (2004). A simple paper model of a transform fault at a spreading ridge. *Journal of Geoscience Education 52*, 391-392.

Ingersoll, E. (1919). Land-Bridges Across the Oceans. In *The Encyclopedia Americana* (Vol. XVI, pp. 692-694). New York, NY: Encyclopedia Americana Corporation. http://en.wikisource.org/wiki/The_Encyclopedia_Americana_(1920)/Land-Bridges_Across_the_Oceans

Kearey, P., & Vine, F. (1996). *Global Tectonics* (2nd E.). Oxford: Blackwell Science Ltd.

Kuenen, P. H. (1937) The negative isostatic anomalies in the East Indies (with Experiments). *Leidse Geologische Mededelingen 8*(2), 169-214. http://www.repository.naturalis.nl/record/505693

Raff, A., & Mason, R. (1961) Magnetic survey off the west coast of North America, 40˚ N to 52˚ N latitude. *Geological Society of America Bulletin 72*, 267-270.

Sinton, J. M., & Detrick, R. S. (1992). Mid-Ocean Ridge Magma Chambers. *Journal of Geophysical Research 97*(B1), 197-216.

Stewart, J. A. (1990). *Drifting continents and colliding paradigms.* Bloomington IN: Indiana University Press.

Tarduno, J. A., Duncan, R. A., Scholl, D. W., Cottrell, R. D., Steinberger, B., Thordarson, T., Kerr, B. C., Neal, C. R., Frey, F. A., Torii, M., & Carvallo, C. (2003). The Emperor Seamounts: Southward Motion of the Hawaiian Hotspot Plume in Earth's Mantle. *Science 301*(5636), 1064–1069. DOI: 10.1126/science.1086442

Thordarson, T., & Larsen, G. (2007) Volcanism in Iceland in historical time: Volcano types, eruption styles and eruptive history. *Journal of Geodynamics 43*, 118–152.

Wegener, A. (1920). *Die Entstehung der Kontinente und Ozeane*. Braunschweig, Germany: Friedr. Vieweg & Sohn. Full text at Project Gutenberg: http://www.gutenberg.org/files/45460/45460-h/45460-h.htm

Wilson, J. T. (1965). A new class of faults and their bearing on continental drift. *Nature 207*, 343-347.

Answers to Review Questions

1. The evidence used by Wegener to support his idea of moving continents included matching continental shapes and geological features on either side of the Atlantic; common terrestrial fossils in South America, Africa, Australia, and India; and data on the rate of separation between Greenland and Europe.
2. The primary weakness of Wegener's model was that he had no realistic mechanism for making continents move.
3. Contractionists believed that mountains formed because the crust wrinkled into mountains as Earth cooled and contracted. Permanentists believed that mountains formed by the geosynclinal process.
4. In the late 19th century the trans-Atlantic paleontological matchups were explained by assuming that there must have been land bridges between the continents at some time in the past.
5. Prior to 1920, ocean depths were measured by dropping a weighted line over the side of ship. Echo sounding techniques were developed at around that time and greatly facilitated the measurement of ocean depths.
6. Paleomagnetic studies showed that old rocks on the continents indicated different locations for magnetic north than the position of magnetic north today. They also showed that the difference in pole position from data on different continents increased progressively for older and older rocks. This implied that either Earth had more than one magnetic pole moving around, or that the continents had moved.
7. Trenches associated with subduction zones are the deepest parts of the oceans.
8. The ocean ridge areas are the youngest parts of the sea floor and thus there hasn't been time for much sediment to accumulate.
9. It was (and still is) thought that high heat flow exists where mantle convection cells are moving hot rock from the lower mantle toward the surface, and that low heat flow exists where there is downward movement of mantle rock.
10. Earthquakes are consistently shallow and relatively small at ocean ridges. At ocean trenches earthquakes become increasingly deep in the direction that the subducting plate is moving.
11. In the Hess model new crust was formed at ocean ridges. Crust was recycled back into the mantle at the trenches.
12. Hess's model did not include the concept of tectonic plates.
13. A mantle plume is a column of hot rock (not magma) that ascends toward the surface from the lower mantle. Mantle plumes last tens of millions of years to hundreds of millions of years.
14. (a) Between the ridge segments there is movement in opposite directions along a transform fault. (b) Outside of the ridge segments the two plates are moving in the same direction and likely at about the same rate. These regions are known as fracture zones.
15. Tectonic plates are made up of crust and the lithospheric (rigid) part of the underlying mantle. The mantle part ensures that the very different oceanic and continental crust sections of a plate can act as one unit.
16. A mantle plume beneath a continent can cause the crust to form a dome that might eventually split open. Several mantle plumes along a line within a continent could lead to rifting.
17. Subduction does not take place at a continent-continent convergent zone because neither plate is dense enough to sink into the mantle.
18. Continental rifting is taking place along the East Africa Rift, and sea floor has recently been created in the Red Sea and also in the Gulf of California.
19. The accumulation of sediment at a passive ocean-continent boundary will lead to the depression of the lithosphere and could eventually result in the separation of the oceanic and continental parts of the plate and the beginning of subduction.

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-NC-SA 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 5. Minerals

Adapted by Karla Panchuk from *Physical Geology* by Steven Earle

Figure 5.1 Giant crystals of gypsum in the Naica Mine in Mexico. The crystals formed in volcanically heated water, and became accessible when the cave was drained as part of mining activities. The cave was very hot, making it fatal for visitors to enter without cooling equipment and respirators. When mining activities ceased, caverns were allowed to flood again. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photograph- Paul Williams (2009) CC BY-NC 2.0 See Appendix C for more attributions.*

Learning Objectives

After reading this chapter and answering the *Review Questions* at the end, you should be able to:

- List the criteria required for a substance to be considered a mineral.
- Explain how atoms bond within minerals.
- Explain how mineral lattices influence the properties of minerals.
- Summarize the categories of minerals defined by anions or anionic groups.
- Describe the types of configurations of silica tetrahedra found in silicate minerals.
- Explain how minerals form.
- Describe the key properties for identifying minerals.

What Is a Mineral?

Minerals are all around us: the graphite in your pencil, the salt on your table, the plaster on your walls, and the trace amounts of gold in your computer. Minerals can be found in a wide variety of consumer products such as paper, medicine, processed foods, and cosmetics. And of course, everything made of metal is also derived from minerals.

A **mineral** is a naturally occurring solid made of specific elements, and arranged in a particular repeating three-dimensional structure.

"Naturally occurring" means that minerals can be formed from substances and under conditions found in nature. Substances that can only be made by humans—classified as **anthropogenic** materials—do not count as minerals, nor do substances produced by natural processes acting upon anthropogenic materials.

In the context of the definition of minerals, "solid" means solid at 25ºC. There are some exceptions to this rule, made for substances defined as minerals before 1959, prior to strict procedures being established for determining what is or isn't a mineral. One example is ice, which is only solid at or below 0 °C. Another is mercury, which is solid below -39 ºC. Mercury that is present in rocks at temperatures above -39 ºC appears as silvery blobs of liquid (Figure 5.2).

Figure 5.2 Droplets of native mercury (pure mercury, Hg), also called quicksilver, amid waxy red crystals of cinnabar (HgS). Cinnabar is a mercury ore mineral. *Source: Parent Géry (2012) CC BY-SA 3.0*

"Specific elements" means that minerals have a specific chemical formula or composition. The mineral pyrite, for example, is FeS_2 (two atoms of sulphur for each atom of iron), and any significant departure from that formula would make it a different mineral. Some minerals can have variable compositions within a specific range. The mineral olivine, for example, has a formula written as $(Fe,Mg)_2SiO_4$, because the composition of olivine can range all the way from Fe_2SiO_4 to Mg_2SiO_4, and have any proportion of iron and magnesium in between. This type of substitution is known as **solid solution**.

Most important of all, the atoms within a mineral are arranged in a specific repeating three-dimensional structure or **lattice**. This regular structure means that all minerals are **crystals**. The mineral **halite**, which we use as table salt, has a relatively simple crystal lattice (Figure 5.3). Atoms of sodium (Na, purple) alternate with atoms of chlorine (Cl, green). The chemical bonds holding the Na and Cl atoms together are all at 90º to each other. Even tiny crystals, like the ones in your salt shaker, have lattices that extend in three dimensions for thousands of repetitions. Halite will always have this structure, and will always have the formula NaCl.

Figure 5.3 Halite crystal lattice. Halite is the mineral in table salt. *Source: Steven Earle (2015) CC BY 4.0*

Some mineral-like materials do not have a regular internal atomic arrangement. Opal (Figure 5.4) is one example. In many respects it fits the definition of a mineral: it has a specific chemical composition ($SiO_2 \cdot nH_2O$, where *n* means that there can be varying amounts of water in the structure), forms naturally through geological processes, and is solid at 25 °C. However, the structure of opal consists of closely packed spheres (Figure 5.4, right) rather than a lattice like halite. Substances like opal, which are mineral-like, but which do not have a crystalline structure, are called **mineraloids**.

Figure 5.4 Opal is mineral-like, but does not have a crystalline structure. Instead, it is made up of layers of closely packed spheres (right). *Source: Left- James St. John (2016) CC BY 2.0; Right- Mineralogy Division, Geological and Planetary Sciences, Caltech (n.d.) CC BY-NC*

Note: Element symbols such as Na and Cl are used extensively in this book. In Appendix 1 you can find a list of the symbols, the names of the elements common in minerals, and a copy of the periodic table of elements.

5.1 Atoms

Protons Are What Make Elements Distinct

All matter, including mineral crystals, is made up of atoms. All atoms are made up of three main particles: **protons**, **neutrons**, and **electrons**. Protons have a positive charge, neutrons have no charge, and electrons have a negative charge. Protons and neutrons have approximately the same mass, but electrons have a mass that is 10,000 times smaller.

The element hydrogen (H) has the simplest atoms. Most hydrogen atoms have just one proton and one electron. The proton forms the **nucleus** (the centre of the atom), while the electron orbits around it (Figure 5.5, left). All other elements have more than one proton in their nucleus. Protons repel each other because they are positively charged, but it is possible to have more than one proton in a nucleus because neutrons hold them together. The next most complex atom, helium (He) has two protons and two neutrons in its nucleus, in its most common form. Some atoms of the same element can have different numbers of neutrons. For example, forms of hydrogen exist with one and two neutrons, and a tiny fraction of He atoms have only one neutron. Forms of an element with different numbers of neutrons are called **isotopes**.

Figure 5.5 Atomic structure of hydrogen and helium showing protons (p+), neutrons (n), and electrons (e-). *Source: Bruce Blaus (2014) CC BY 3.0*

The number of protons in an atom determines what element it will be, so the number of protons is called the **atomic number** of that element. The total number of protons and neutrons in the nucleus is the **mass number**. The mass number distinguishes between isotopes of an element. Isotopes of an element are denoted by putting the mass number as a subscript in front of the symbol for that element. For example, the isotopes of hydrogen are ^{1}H (1 proton), ^{2}H (1 proton + 1 neutron), and ^{3}H (1 proton + 2 neutrons).

For most of the 16 lightest elements (up to oxygen) the number of neutrons is equal to the number of protons. For most of the remaining elements, there are more neutrons than protons. This is because the more protons that are concentrated in a small space, the more neutrons are needed to keep the nucleus together. The most common isotope of uranium (U), for example, is ^{238}U. It has 92 protons, but requires 146 neutrons to keep them together. The neutrons are only partly successful. Uranium is radioactive, meaning that its nucleus will eventually split apart and release energy. What remains of the nucleus has fewer protons, so after decay the atom is a different element.

Electrons Are What Control How Atoms Interact

Electrons orbiting around the nucleus of an atom are arranged in shells (also called **energy levels**). The first shell can hold only two electrons (as in H and He in Figure 5.5), but the next shell holds up to eight electrons. An atom can have many shells of electrons, but there are never more than 8 outermost electrons interacting with surrounding atoms.

The outermost electrons determine how atoms can be bonded together. Elements that have a full outer shell (e.g., neon, Figure 5.6 right) are **inert** because they do not react with other elements to form compounds. These are the **noble gases** (including helium, argon, krypton, and radon, in addition to neon) in the far-right column of the periodic table. For elements that do not have a full outer shell (e.g., lithium, Figure 5.6 left), the outermost electrons can interact with the outermost electrons of nearby atoms to create chemical bonds.

Figure 5.6 The number of electrons in an atom's outermost shell (or energy level) determine whether it will bond to other atoms, and how it will bond. Right- Neon has a completely filled outer shell with 8 electrons. It does not bond with other atoms. Left- Lithium has only one electron in its outer shell. It bonds with other atoms. *Source: Bruce Blaus (2014) CC BY 3.0*

The electron shell configurations for 29 of the first 36 elements are listed in Table 5.1. Note that some of the shells in the table below have more than 8 electrons. This is because they contain subshells. For

example, the third shell can hold up to 18 electrons because it contains one subshell that can hold 2 electrons, and two subshells that can hold 8 electrons each.

Table 5.1 Electron shell configurations of some of the elements up to krypton. Inert elements (those with filled outer shells) are shaded.

Element	Symbol	Atomic Number	Number of electrons in each shell			
			First	Second	Third	Fourth
Hydrogen	H	1	1			
Helium	**He**	**2**	**2**			
Lithium	Li	3	2	1		
Beryllium	Be	4	2	2		
Boron	B	5	2	3		
Carbon	C	6	2	4		
Nitrogen	N	7	2	5		
Oxygen	O	8	2	6		
Fluorine	F	9	2	7		
Neon	**Ne**	**10**	**2**	**8**		
Sodium	Na	11	2	8	1	
Magnesium	Mg	12	2	8	2	
Aluminum	Al	13	2	8	3	
Silicon	Si	14	2	8	4	
Phosphorus	P	15	2	8	5	
Sulphur	S	16	2	8	6	
Chlorine	Cl	17	2	8	7	
Argon	**Ar**	**18**	**2**	**8**	**8**	
Potassium	K	19	2	8	8	1
Calcium	Ca	20	2	8	8	2
Scandium	Sc	21	2	8	9	2
Titanium	Ti	22	2	8	10	2
Vanadium	V	23	2	8	11	2
Chromium	Cr	24	2	8	13	1
Manganese	Mn	25	2	8	13	2
Iron	Fe	26	2	8	14	2
-						
Selenium	Se	34	2	8	18	6
Bromine	Br	35	2	8	18	7
Krypton	**Kr**	**36**	**2**	**8**	**18**	**8**

5.2 Bonding and Lattices

Atoms seek to have a full outer shell. For hydrogen and helium, a full outer shell means two electons. For other elements, it means 8 electrons. Filling the outer shell is accomplished by transferring or sharing electrons with other atoms in chemical bonds. The type of chemical bond is important for the study of minerals because the type of bond will determine many of a mineral's physical and chemical properties.

Ionic Bonds

Consider the example of halite again, which is made up of sodium (Na) and chlorine (Cl). Na has 11 electrons: two in the first shell, eight in the second, and one in the third (Figure 5.7, top). Na readily gives up the third shell electron so it can have the second shell with 8 electrons as its outermost shell. When it loses the electron, the total charge from the electrons is -10, but the total charge from the protons is +11, so it is left with a +1 charge over all.

Figure 5.7 Electron configuration of sodium and chlorine atoms (top). Sodium gives up an electron to become a cation (bottom left) and chlorine accepts an electron to become an anion (bottom right). *Source: Steven Earle (2015) CC BY 4.0*

Chlorine has 17 electrons: two in the first shell, eight in the second, and seven in the third. Cl readily accepts an eighth electron to fill its third shell, and therefore becomes negatively charged because it has a total charge of -18 from electrons, and a total charge of +17 from protons.

In changing their number of electrons, these atoms become **ions** — the sodium loses an electron to become a positive ion or **cation**[1], and the chlorine gains an electron to become a negative ion or **anion** (Figure 5.7, bottom). Because negative and positive charges attract, sodium and chlorine ions stick together, creating an **ionic bond**. In an ionic bond, electrons can be thought of as having transferred from one atom to another.

[1] You can remember that a cation is positive by remembering that a cat has paws (paws sounds like "pos" in "positive"). You could also think of the "t" in "cation" as a plus sign.

Exercise: Cation or Anion?

A number of elements are listed below along with their atomic numbers (the number of protons, and therefore also the number of electrons in the atom). Assuming that the first electron shell can hold two electrons and subsequent electron shells can hold eight electrons, sketch the electron configurations for these elements, as in the example for fluorine (Figure 5.8). If you fill a shell and have electrons left over, draw another shell around the atom. Predict whether the element is likely to form a cation or an anion, and what charge it would have (e.g., +1, +2, –1).

1. Lithium (3)
2. Magnesium (12)
3. Argon (18)
4. Chlorine (17)
5. Beryllium (4)
6. Oxygen (8)
7. Sodium (11)

Figure 5.8 How to draw the electron configuration for fluorine, with an atomic number of 9. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0*

Covalent Bonds

An element like chlorine can also form bonds without forming ions. For example, two chlorine atoms can each complete their outer shells by sharing electrons. Chlorine gas (Cl_2, Figure 5.9) is formed when two chlorine atoms form a **covalent bond**.

Figure 5.9 A covalent bond between two chlorine atoms. The electrons are black in the left atom, and blue in the right atom. Two electrons are shared (one black and one blue) so that each atom appears to have a full outer shell. *Source: Steven Earle (2015) CC BY 4.0*

Carbon is another atom that participates in covalent bonding. An uncharged carbon atom has six protons and six electrons. Two of the electrons are in the inner shell and four are in the outer shell (Figure 5.10, left). Carbon would need to gain or lose four electrons to have a filled outer shell, and this would create too great a charge imbalance. Instead, carbon atoms share electrons to create covalent bonds (Figure 5.10, right).

Figure 5.10 The electron configuration of carbon (left) and the sharing of electrons in covalent C bonding (right). The electrons shown in blue are shared between adjacent C atoms. *Source: Steven Earle (2015) CC BY 4.0*

In the mineral diamond (Figure 5.11, left), the carbon atoms are linked together in a three-dimensional framework, where one carbon atom is bonded to four other carbon atoms, and every bond is a very strong covalent bond.

Figure 5.11 Covalently bonded structures. Left: Diamond with three-dimensional structure of covalently bonded carbon. Right: Graphite with covalently bonded sheets of carbon. Sheets are held together by weaker van der Waals forces. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Materialscientist (2009) CC BY-SA 3.0*

Other Types of Bonds

Most minerals are characterized by ionic bonds, covalent bonds, or a combination of the two, but there are other types of bonds that are important in minerals. Consider the mineral graphite (Figure 5.11, right): the carbon atoms are linked together in sheets or layers in which each carbon atom is covalently bonded to three others. Graphite-based compounds are strong because of the covalent bonding between carbon atoms within each layer, which is why they are used in high-end sports equipment such as ultralight racing bicycles. Graphite itself is soft, however, because the layers themselves are held together by relatively weak **Van der Waals forces**.

Van der Waals forces, like **hydrogen bonds**, work because molecules can be electrostatically neutral, but still have an end that is slightly more positive and an end that is slightly more negative. In water molecules (Figure 5.12, left), the bent shape puts the hydrogen atoms on one side of the molecule, and the oxygen atom, with more electrons, on the other. The charge is distributed asymmetrically across the water molecule. Contrast this with the straight carbon dioxide molecule (Figure 5.12, right). The slightly more negative oxygen atoms on the ends are distributed symmetrically on either side of the carbon atom.

Figure 5.12 Hydrogen bonding. Water molecules (left) are polar molecules (their charge is distributed asymmetrically). Slightly negative parts of the molecule are attracted to slightly positive parts of other water molecules. CO_2 (right) is a non-polar molecule. The slightly negative oxygen atoms are distributed symmetrically on either side of the carbon atom. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Modified after Querter (2011) CC BY-SA 3.0 and Jynto (2011) CC0 1.0*

Metallic bonding occurs in metallic elements because they have outer electrons that are relatively loosely held. (The metals are highlighted on the periodic table in Appendix 1.) When bonds between such atoms are formed, the **dissociated electrons** can move freely from one atom to another. This feature accounts for two very important properties of metals: their electrical conductivity and their malleability (they can be deformed and shaped).

Figure 5.13 Metallic bonding. Dissociated electrons (grey dots) move between metal atoms. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Nucleus by Fornax (2010) CC BY-SA 3.0*

5.3 Mineral Groups

Minerals are organized according to the anion or anion group (a group of atoms with a net negative charge, e.g., SO_4^{2-}) they contain, because the anion or anion group has the biggest effect on the properties of the mineral. Silicates, with the anion group SiO_4^{4-}, are by far the most abundant group in the crust and mantle. (They will be discussed in Section 5.4). The different mineral groups along with some examples of minerals in each group are summarized below.

Oxide Minerals: O^{2-} Anion

Oxide minerals (Figure 5.14) have oxygen (O^{2-}) as their anion. They don't include anion groups with other elements, such as the carbonate (CO_3^{2-}), sulphate (SO_4^{2-}), and silicate (SiO_4^{4-}) anion groups. The iron oxides hematite and magnetite are two examples that are important ores of iron. Corundum is an abrasive, but can also be a gemstone in its ruby and sapphire varieties. If the oxygen is also combined with hydrogen to form the hydroxyl anion (OH^-), the mineral is known as a **hydroxide**. Some important hydroxides are limonite and bauxite, which are ores of iron and aluminum, respectively.

Figure 5.14 Oxide minerals include metal ore minerals, industrial minerals, and gemstones. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photos by R. Weller/ Cochise College.*

Sulphide Minerals: S^{2-} Anion

Sulphide minerals (Figure 5.15) include galena, sphalerite, chalcopyrite, and molybdenite, which are the most important ores of lead, zinc, copper, and molybdenum, respectively. Some other sulphide minerals are pyrite, bornite, stibnite, and arsenopyrite. Sulphide minerals tend to have a metallic sheen.

Figure 5.15 Sulphide minerals often have a metallic lustre and include metal ores. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photos by R. Weller/ Cochise College.*

Sulphate Minerals: SO_4^{2-} Anion Group

Many **sulphate** minerals form when sulphate-bearing water evaporates. A deposit of sulphate minerals may indicate that a lake or sea has dried up at that location. Sulphates with calcium include anhydrite, and gypsum (Figure 5.16). Sulphates with barium and strontium are barite and celestite, respectively. In all of these minerals, the cation has a +2 charge, which balances the –2 charge on the sulphate ion.

Figure 5.16 Sulphate minerals. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0*

Halide Minerals: Anions from the Halogen Group

The anions in **halides** are the **halogen** elements including chlorine, fluorine, and bromine. Examples of halide minerals are cryolite, fluorite, and halite (Figure 5.17). Halide minerals are made of ionic bonds. Like the sulphates, some halides also form when mineral-rich water evaporates.

Figure 5.17 Halide minerals. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0.*

Carbonate Minerals: CO_3^{2-} Anion Group

The **carbonate** anion group combines with +2 cations to form minerals such as calcite, magnesite, dolomite, and siderite (Figure 5.18). The copper minerals malachite and azurite are also carbonates. The carbonate mineral calcite is the main component of rocks formed in ancient seas by organisms such as corals and algae.

Phosphate Minerals: PO_4^{3-} Anion

The apatite group of **phosphate** minerals (Figure 5.19, left) includes hydroxyapatite, which makes up the enamel of your teeth. Turquoise is also a phosphate mineral (Figure 5.19, right).

Native Element Minerals

These are minerals made of a single element, such as gold, copper, silver, or sulphur (Figure 5.20).

Figure 5.18 Carbonate minerals. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Photos by Rob Lavinsky, iRocks.com, CC BY-SA 3.0.*

Figure 5.19 Phosphate minerals. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photos by R. Weller/ Cochise College.*

Figure 5.20 Native element minerals are made up of a single element. *Source: Karla Panchuk (2018) CC BY-SA 4.0.*

Exercise: Mineral Groups

Minerals are grouped according to the anion part of the mineral formula, and mineral formulas are always written with the anion part last. For example, for pyrite (FeS_2), Fe^{2+} is the cation and S^{-} is the anion. This helps us to know that it's a sulphide, but it is not always that obvious. Hematite (Fe_2O_3) is an oxide; that's easy, but anhydrite ($CaSO_4$) is a sulphate because SO_4^{2-} is the anion, not O. Similarly, calcite ($CaCO_3$) is a carbonate, and olivine (Mg_2SiO_4) is a silicate. Minerals with only one element (such as S) are native minerals, while those with an anion from the halogen column of the periodic table (Cl, F, Br, etc.) are halides.

Provide group names for the following minerals:

Mineral	Formula	Mineral	Formula
Sphalerite	ZnS	Silver	Ag
Magnetite	Fe_3O_4	Fluorite	CaF_2
Pyroxene	$MgSiO_3$	Ilmenite	$FeTiO_3$
Anglesite	$PbSO_4$	Siderite	$FeCO_3$
Sylvite	KCl	Feldspar	$KAlSi_3O_8$
Sulfur	S	Xenotime	YPO_4

5.4 Silicate Minerals

Silicon and oxygen bond covalently to create a **silicate tetrahedron** (SiO_4^{4-}), which is a four-sided pyramid shape with oxygen at each corner and silicon in the middle (Figure 5.21). This structure is the building block of many important minerals in the crust and mantle. Silicon has a charge of +4, and oxygen has a charge of -2, so the total charge of the silicate anion is -4.

Figure 5.21 The silica tetrahedron is the building block of all silicate minerals. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Modified after Helgi (2013) CC BY-SA 3.0*

In silicate minerals, these tetrahedra are arranged and linked together in a variety of ways, from single units to chains, rings, and more complex frameworks. In the rest of this section we will discuss the structures of the most common silicate minerals in Earth's crust and mantle.

Isolated Tetrahedra

The simplest silicate structure, that of the mineral **olivine** (Figure 5.23), is composed of isolated tetrahedra bonded to iron and/or magnesium ions (Figure 5.23 left). In olivine, the –4 charge of each silica tetrahedron is balanced by two iron or magnesium cations, each with a charge of +2. Olivine can be pure Mg_2SiO_4 or Fe_2SiO_4, or a combination of the two, written as $(Mg,Fe)_2SiO_4$. Magnesium and iron can substitute for each other because they both have a charge of +2, and they are similar in size. Magnesium cations have a radius of 0.73 Å, and iron cations have a radius of 0.62 Å[2].

[2] Å stands for Ångstrom, a unit commonly used to express atomic-scale dimensions. One angstrom is 10^{-10} m or 0.0000000001 m

Figure 5.23 Olivine is a silicate mineral made of isolated silica tetrahedra bonded to Fe and Mg ions (left). Olivine crystals (centre) can often be found in the volcanic igneous rock called basalt (right). *Source: Karla Panchuk (2018) CC BY-SA 4.0. Left- modified after Steven Earle (2015) CC BY 4.0*

Although the iron and magnesium ions are similar in size, allowing them to substitute for each other in some silicate minerals, the common ions in silicate minerals have a wide range of sizes (Figure 5.24). Ionic radii are critical to the composition of silicate minerals, because the structure of the silicate mineral will determine the size of spaces available.

Figure 5.24 The ionic radii in angstroms of some of the common ions in silicate minerals. Radii shown to scale. Notice that iron appears twice with two different radii. This is because iron can exist as a +2 ion (if it loses two electrons when it becomes an ion) or a +3 ion (if it loses three). Fe^{2+} is known as ferrous iron. Fe^{3+} is known as ferric iron. *Source: Karla Panchuk (2017) CC BY 4.0. Modified after Steven Earle (2015) CC BY 4.0*

Chain Silicates

Pyroxene (Figure 5.25 bottom left) is an example of a single-chain silicate. The structure of chain silicates is shown in Figure 5.25 (top). In pyroxene, silica tetrahedra form a chain because one oxygen from each tetrahedron is shared with the adjacent tetrahedron. This means there are fewer oxygens in the structure. This can be expressed as an oxygen-to-silicon ratio (O:Si). The O:Si is lower than in olivine (3:1 instead of 4:1), and the net charge per silicon atom is less (–2 instead of –4), because fewer cations are necessary to balance that charge.

Pyroxene compositions have the silica tetrahedra represented as SiO_3 (e.g., $MgSiO_3$, $FeSiO_3$, and $CaSiO_3$.[3] In other words, pyroxene has one cation for each silica tetrahedron (e.g., $MgSiO_3$) while olivine has two (e.g., Mg_2SiO_4). The structure of pyroxene is more "permissive" than that of olivine, meaning cations with a wider range of ionic radii can fit into it. That's why pyroxenes can have calcium cations (radius 1.00 Å) substitute for iron (0.63 Å) and magnesium (0.72 Å).

[3] The variation in composition can also be written as $(Mg,Fe,Ca)SiO_3$, where the elements in the brackets can be present in any proportion

Figure 5.25 Chain silicate minerals. Top: Arrangement of silica tetrahedra in single and double chains. Bottom left: Pyroxene crystals (dark crystals) of the variety aegirine (acmite). Bottom right: Amphibole crystal (dark) of the variety hornblende. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Top left- modified after Steven Earle (2015) CC BY 4.0. Top right- modified after Klein & Hurlbut (1993). Photos by R. Weller/ Cochise College.*

In **amphibole** (Figure 5.25 bottom right), the silica tetrahedra are linked in a double chain that has an oxygen-to-silicon ratio lower than that of pyroxene, and hence still fewer cations are necessary to balance the charge. Amphibole is even more permissive than pyroxene and its compositions can be very complex, as shown by the formula for the hornblende group of amphibole minerals in Figure 5.25 (bottom right).

Exercise: Oxygen to Silicon Ratio

Figure 5.26 shows single chain and double chain structures. Count the number of tetrahedra versus the number of oxygen ions (yellow spheres) for each. Each tetrahedron has one silicon atom.

1. Confirm for yourself that the ratio of silicon to oxygen in the single chain is 1:3.
2. What is the O:Si for the double chain?

Figure 5.26 Single and double chains of tetrahedra. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0*

Sheet Silicates

In **mica** structures the silica tetrahedra are arranged in continuous sheets (Figure 5.27), where each tetrahedron shares three oxygen anions with adjacent tetrahedra. Because even more oxygens are shared between adjacent tetrahedra, fewer charge-balancing cations are needed for sheet silicate minerals. Bonding between sheets is relatively weak, and this accounts for the tendency of mica minerals to split apart in sheets (Figure 5.27 bottom right). Two common micas in silicate rocks are **biotite** (Figure 5.27 bottom left), which contains iron and/or magnesium, making it a dark mineral; and muscovite (Figure 5.27 right), which contains aluminum and potassium, and is light in colour. All of the sheet silicate minerals have water in their structure, in the form of the hydroxyl (OH-) anion.

Figure 5.27 Micas are sheet silicates and split easily into thin layers along planes parallel to the sheets. Biotite mica (lower left) is has Fe and Mg cations. Muscovite mica (lower right) has Al and K instead. The muscovite mica shows how thin layers can split away in a sheet silicate. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Top left- modified after Steven Earle (2015) CC BY 4.0. Top right- modified after Klein & Hurlbut (1993). Photos by R. Weller/ Cochise College.*

Some sheet silicates typically occur in clay-sized fragments (i.e., less than 0.004 mm). These include the clay minerals **kaolinite**, **illite,** and **smectite**, which are important components of rocks and especially of soils.

Framework Silicates

In framework silicates, tetrahedra are connected to each other in three-dimensional structures rather than in two-dimensional chains and sheets.

Feldspar

Feldspars are a group of very abundant framework silicates in Earth's crust. They include alumina tetrahedra as well as silicate tetrahedra. In alumina tetrahedra, there is an aluminum cation at the centre instead of a silicon cation.

Feldspars are classified using a ternary (3-fold) system with three end-members ("pure" feldspars). This system is illustrated with a triangular diagram that has each end-member at one corner (Figure 5.28). The distance along a side of the diagram represents the relative abundance of the composition of each end-member.

Figure 5.28 Ternary diagram showing the feldspar group of framework silicate minerals. Alkali feldspars are those with compositions ranging between albite (with a Na cation) and orthoclase and its polymorphs (with a K cation. Plagioclase feldspars are those with compositions ranging between albite and anorthite (with a Ca cation). *Source: Karla Panchuk (2018) CC BY-SA 4.0. Ternary diagram modified after Klein & Hurlbut (1993).*

One end-member is **potassium feldspar** (also referred to as **K-feldspar**), which has the composition $KAlSi_3O_8$. Depending on the temperature and rate of cooling, K-feldspar can occur as one of three polymorphs: **orthoclase**, **sanidine**, or **microcline**. Another end member is **albite**, which has sodium instead of potassium (formula $NaAlSi_3O_8$). As is the case for iron and magnesium in olivine, there is a continuous range of compositions (referred to as a solid-solution series) between albite and orthoclase. Feldspars in this series are referred to as **alkali feldspars**. Potassium cations are much larger than sodium cations (1.37 Å versus 0.99 Å, respectively), so high temperatures are required to form alkali feldspars with intermediate compositions.

The third end-member is **anorthite** and it has calcium instead of potassium or sodium (formula $CaAl_2Si_2O_8$). Feldspars in the solid-solution series between albite and anorthite are called **plagioclase feldspars**. Calcium and sodium cations are nearly the same size (1.00 Å and 0.99 Å, respectively), so from that perspective it makes sense that they substitute readily for each other, and that any intermediate compositions between $CaAl_2Si_2O_8$ and $NaAlSi_3O_8$ can exist. However, calcium and sodium ions don't have the same charge (Ca^{2+} versus Na+), making it surprising that they substitute so easily. The difference in charge is accommodated by substituting some Al^{3+} for Si^{4+}. Albite has one Al and three Si in its formula, while anorthite is has two Al and two Si. Plagioclase feldspars of intermediate composition also have intermediate proportions of Al and Si.

Quartz

Quartz (SiO_2; Figure 5.29) contains only silica tetrahedra. In quartz, each silica tetrahedron is bonded to four other tetrahedra (with an oxygen shared at every corner of each tetrahedron), making a three-dimensional framework. As a result, the ratio of silicon to oxygen is 1:2. Because the one silicon cation has a +4 charge and the two oxygen anions each have a –2 charge, the charge is balanced. There is no need

to add cations to balance the charge. The hardness of quartz and the fact that it breaks irregularly (notice the bottom of the crystal in Figure 5.29 right) and not along smooth planes result from the strong covalent/ionic bonds characteristic of the silica tetrahedron.

Figure 5.29 Quartz is another silicate mineral with a three-dimensional framework of silica tetrahedra. Sometimes quartz occurs as well-developed crystals (left), but it also occurs in common rocks such as granite (right). In addition to quartz, the granite contains potassium feldspar, albite, and amphibole. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photos by R. Weller/ Cochise College.*

5.5 How Minerals Form

The following criteria are required for mineral crystals to grow:

- The elements needed to make the mineral crystals must be present in sufficient abundance and appropriate proportions.
- The physical and chemical conditions must be favourable.
- There must be sufficient time for the atoms to become arranged into a lattice.

Physical and chemical conditions include factors such as temperature, pressure, amount of oxygen available, pH, and the presence of water. The presence of water makes it easier for ions to move to where there are needed, and can lead to the formation of larger crystals over shorter time periods, as with the gypsum crystals at the beginning of this chapter. Time is one of the most important factors because it takes time for atoms to line themselves up into an orderly structure. If time is limited, the mineral grains may remain very small.

Most of the minerals that make up the rocks in the crust and mantle formed through the cooling of molten rock, known as **magma**. At the high temperatures that exist deep within Earth, some geological materials are liquid. As magma rises up through the crust, either by volcanic eruption or by more gradual processes, it cools and minerals crystallize. When cooling is rapid and many crystals form at once, only small mineral grains will form before the rock becomes solid. The resulting rock will be fine-grained (i.e., crystals less than 1 mm). When cooling is slow, or when few crystals are growing at a time, relatively large crystals will develop.

Minerals can also form in several other ways:

- Precipitation from a solution (e.g., from hot water flowing underground, or when evaporation concentrates ions in a lake or inland sea)
- Precipitation from a gas (e.g., from vents releasing volcanic gases)
- Metamorphism: Solid minerals react with each other under high pressures and temperatures, and new minerals are formed.
- Weathering: Minerals unstable at Earth's surface are chemically altered by surface processes.
- Organic formation: Organisms build shells (primarily of calcite or aragonite), and teeth and bones (primarily of apatite).

5.6 Mineral Properties

Minerals are universal. A crystal of hematite on Mars will have the same properties as one on Earth, and the same as one on a planet orbiting another star. That's good news for geology students who are planning interplanetary travel, because they can use the same properties to identify minerals anywhere. That doesn't

mean that it's easy, however. Identification of minerals takes practice. Some of the mineral properties that are useful for identification are colour, streak, lustre, hardness, habit, cleavage or fracture, and density.

Colour

Some minerals have distinctive colours that useful as diagnostic criteria. The mineral sulphur (Figure 5.30 left) is always a characteristic bright yellow. For other minerals, colour might vary. Hematite is an example of a mineral for which colour is not necessarily diagnostic. In some forms hematite is a deep dull red (a fairly unique colour), but in others it is a metallic silvery black (5.30, right).

Figure 5.30 Colour is a useful diagnostic property for sulphur (left) and for some types of hematite (right) because the yellow and dark red colours are unique to those minerals. In contrast, silvery metallic forms of hematite are similar in appearance to many other minerals. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photos by R. Weller/ Cochise College.*

For other minerals, the problem is that a single mineral can have a wide range of colours. The colour variations can be the result of varying proportions of trace elements within the mineral, or structural defects within the crystal lattice. In the case of quartz (Figure 5.31), milky quartz gets its white colour from millions of tiny fluid-filled cavities. Smoky quartz gets its grey colour from structural damage caused by natural radiation. Amethyst and citrine get their colours from trace amounts of iron, and rose quartz gets its pink hue from manganese.

Streak

The colour of a mineral is what you see when light reflects off the surface of the sample. One reason that colour can be so variable is that the surface texture is variable. A way to get around this problem is to grind a small amount of the sample to a powder and observe the colour of the powder. This colour is the mineral's **streak**. The mineral can be powdered by scraping the sample across a piece of unglazed porcelain called a **streak plate** (Figure 5.32). In Figure 5.32, two samples of hematite have been scraped across the streak plate. Even though one sample is metallic and the other is deep red, both have a similar reddish-brown streak.

Figure 5.31 The many colours of quartz.Quartz can be colourless, milky, a greyish smoky colour, purple, yellow, and pink. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photos by R. Weller/ Cochise College.*

Streak is an especially helpful property when minerals look similar. In Figure 5.33 all of the minerals are dark in colour, with varying degrees of metallic sheen. The streaks of these minerals are much more distinctive.

TOP- **Figure 5.32** Hematite leaves a distinctive reddish-brown streak whether the sample is metallic or deep red. *Source: Karla Panchuk (2015) CC BY 4.0*

RIGHT- **Figure 5.33** Similar dark-grey minerals with varying degrees of metallic sheen leave different colours of streaks. The minerals are from upper left clockwise: hematite, magnetite, sphalerite, and galena. *Source: Karla Panchuk (2015) CC BY 4.0*

Lustre

Lustre is the way light reflects off the surface of a mineral, and the degree to which it penetrates into the interior. The key distinction is between **metallic** and **non-metallic lustre**. Light does not pass through metals, and that is the main reason they look metallic (e.g., the hematite on left of Figure 5.32). Even a thin sheet of metal — such as aluminum foil — will be not permit light to pass through it. Many non-metallic minerals may look as if light will not pass through them, but if you look closely at a thin slice of the mineral you will see that the mineral is translucent or transparent.

If a non-metallic mineral has a shiny, reflective surface, it is said to have a **glassy** lustre. The quartz crystals in Figure 5.31 are examples of minerals with glassy lustre. If the mineral surface is dull and non-reflective, it has an **earthy** lustre (like the hematite on the right of Figure 5.32). Other types of non-metallic lustres are silky, pearly, and resinous. Lustre is a good diagnostic property because most minerals will always appear either metallic or non-metallic, although as Figure 5.31 shows, there are exceptions.

Hardness

One of the most important diagnostic properties of a mineral is its hardness. In practical terms, hardness determines whether or not a mineral can be scratched by a particular material.

In 1812 German mineralogist Friedrich Mohs came up with a list of 10 minerals representing a wide range of hardness, and numbered them 1 through 10 in order of increasing hardness (Figure 5.34, horizontal axis). While each mineral on the list is harder than the one before it, the measured hardness (vertical axis) is not linear. Notice that apatite is about three times harder than fluorite, and diamond is three times harder than corundum.

Figure 5.34 Minerals and reference materials in the Mohs scale of hardness. The measured hardness values are Vickers Hardness numbers. *Source: Steven Earle (2015) CC BY 4.0*

Some commonly available reference materials are also shown on the diagram, including a typical fingernail (2.5)[4], a piece of copper wire (3.5), a knife blade or piece of window glass (5.5), a hardened steel file (6.5), and a porcelain streak plate (7). These are tools that a geologist can use to measure the hardness of unknown minerals: if you have a mineral that you can't scratch with your fingernail, but you can scratch with a copper wire, then its hardness is between 2.5 and 3.5. The minerals themselves can be used to test other minerals.

Crystal Habit

When minerals form within rocks, there is a possibility that they will form in distinctive crystal shapes if they are not crowded out by other pre-existing minerals. Every mineral has one or more distinctive **crystal**

[4] Note that artificial fingernails may be much harder than natural fingernails. Some materials used for artificial nails are harder than quartz.

habits determined by their atomic structure, although it is not that common in ordinary rocks for the shapes to be obvious.

Quartz, for instance, will form six-sided prisms with pointed ends (Figure 5.35 left), but this typically happens only when it crystallizes from a hot water solution within a cavity in an existing rock. Pyrite can form cubic crystals (Figure 5.35 centre), but can also form crystals with 12 faces, known as dodecahedra. The mineral garnet also forms many-sided crystals with an over-all rounded shape (Figure 5.35 right).

Figure 5.35 Hexagonal prisms of quartz (left), intergrown cubic crystals of pyrite (centre), and 24-sided crystals of garnet (right). *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photos by R. Weller/ Cochise College.*

Some of the terms that are used to describe habit include bladed, botryoidal (grape-like), dendritic (branched), drusy (an encrustation of crystals), equant (similar size in all dimensions), fibrous, platy, prismatic (long and thin), and stubby.

Cleavage and Fracture

Cleavage and fracture describe how a mineral breaks. These characteristics are the most important diagnostic features of many minerals, and often the most difficult to understand and identify. **Cleavage** is what we see when a mineral breaks along a plane or planes, while **fracture** is an irregular break. Some minerals tend to cleave along planes at various fixed orientations. Some, like quartz, do not cleave at all, only fracture. Minerals that have cleavage can also fracture along surfaces that are not parallel to their cleavage planes.

The way minerals break is determined by the arrangement of atoms within them, and more specifically by the orientation of weaknesses within their crystal lattice. Graphite and mica break off in parallel sheets (Figure 5.36).

Figure 5.36 One direction of cleavage (basal cleavage). Left: Schematic of basal cleavage. Right: Muscovite showing basal cleavage. The white dashed line marks the edge of the cleavage plane. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Cleavage diagram modified after M.C. Rygel (2010) CC BY-SA 3.0*

Other minerals have two directions of cleavage, classified as two directions at 90° (Figure 5.37 top) and two directions not at 90° (Figure 5.37 bottom). While the diagrams of planes on the left of Figure 5.37 make this difference clear, it may be less obvious in practice. The minerals in Figure 5.37 both have two planes of cleavage that are very close to 90°. The white dashed lines mark the edges of the planes, as with Figure 5.36. See if you can find the planes repeated in the images. The images are close-up views of the minerals, only a few cm across. Sometimes you must look very carefully to find cleavage planes.

Figure 5.37 Two directions of cleavage. Top: Two directions at 90° in pyroxene. Bottom: two directions not at 90° degrees in plagioclase feldspar. Edges of cleavage planes marked with dashed lines. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Cleavage diagrams modified after M.C. Rygel (2010) CC BY-SA 3.0*

Some minerals have many directions of cleavage. Figure 5.38 shows minerals with three directions of cleavage. Halite (Figure 5.38 top) has three directions at 90° and calcite (Figure 5.38 bottom) has three directions not at 90°.

Figure 5.38 Three directions of cleavage. Top: Three directions at 90° in halite. Bottom: Three directions not at 90° in calcite. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Cleavage diagrams modified after M.C. Rygel (2010) CC BY-SA 3.0*

There are a few common difficulties that students encounter when learning to recognize and describe cleavage. One is that it might be necessary to look very closely at a sample to see mineral cleavage. The key features in Figure 5.37 are only cm or mm in scale. If crystals are very small, it may not be possible to see cleavage at all. Another issue is that sometimes cleavage is present, but it is poor, meaning the cleavage surface isn't perfectly flat. Finally it can be difficult to know whether a flat surface on a crystal is a cleavage plane, a crystal face, or simply a surface that happens to be flat. Cleavage planes tend to repeat themselves at different depths throughout the mineral, so if you are unsure whether the surface you are looking at is a cleavage plane, try rotating the mineral in bright light. If cleavage is present, you will generally find that, for a given cleavage direction, all of the cleavage surfaces will glint in the light simultaneously. Crystal faces will also glint in light, but they do not repeat themselves at depth throughout the mineral. The best way to overcome all of these problems is to look at lots of examples. It's worth it to be able to identify cleavage and fracture, because cleavage is a reliable diagnostic property for most minerals.

Density

Density is a measure of the mass of a mineral per unit volume, and it is a useful diagnostic tool in some cases. Most common minerals, such as quartz, feldspar, calcite, amphibole, and mica, are of average density (2.6 to 3.0 g/cm^3), and it would be difficult to tell them apart on the basis of their density. On the other hand, many of the metallic minerals, such as pyrite, hematite, and magnetite, have densities over 5 g/cm^3. If you picked up a sample of one of these minerals, it would feel much heavier compared to a similarly sized sample of a mineral with average density. A limitation of using density as a diagnostic tool is that one cannot assess it in minerals that are a small part of a rock with other minerals in it.

Other properties

Several other properties are useful for identification of some minerals. Some of these are:

- Calcite reacts with dilute acid and will give off bubbles of carbon dioxide.
- Magnetite is strongly magnetic, and some other minerals are weakly magnetic.
- Sphalerite ((Zn,Fe)S) gives off a smell of sulphur when drawn across a streak plate.
- Halite tastes salty.
- Talc feels soapy to the touch.
- Plagioclase feldspar has striations (parallel razor-thin lines etched on the surface) and some varieties show a play of colours when light hits them at the right angle (see the labradorite in Figure 5.28).

Chapter 5 Summary

The topics covered in this chapter can be summarized as follows:

5.1 Atoms

An atom is made up of protons and neutrons in the nucleus, and electrons arranged in energy shells around the nucleus. The first shell holds two electrons, and outer shells hold more. Atoms strive to have eight electrons in their outermost shell (or two for H and He). Atoms gain, lose, or share electrons to achieve this. In so doing they become either positively charged cations (if they lose electrons) or negatively charged anions (if they gain them).

5.2 Bonding and Lattices

The main types of bonding in minerals are ionic bonding (electrons transferred) and covalent bonding (electrons shared). Some minerals have metallic bonding or weak Van der Waals forces. Minerals form in three-dimensional lattices. The configuration of the lattices and the type of bonding within help determine mineral properties.

5.3 Mineral Groups

Minerals are grouped according to the anion part of their formula. Some common types are: oxides, sulphides, sulphates, halides, carbonates, phosphates, silicates, and native minerals.

5.4 Silicate Minerals

Silicate minerals are the most common minerals in Earth's crust and mantle. They all have silica tetrahedra (four oxygens surrounding a single silicon atom) arranged in different structures (chains, sheets, etc).

5.5 How Minerals Form

Most minerals in the crust form from the cooling and crystallization of magma. Some form from hot water solutions, during metamorphism or weathering, or through organic processes. More rarely, minerals precipitate directly from a gas, such as at a volcanic vent.

5.6 Mineral Properties

Some of the important properties for mineral identification include hardness, cleavage/fracture, density, lustre, colour, and streak colour.

Review Questions

1. What is the electrical charge of a proton? A neutron? An electron? What are their relative masses?

2. Explain how the need for an atom's outer shell to be filled with electrons contributes to bonding.

3. Why are helium and neon non-reactive?

4. What is the difference in the role of electrons in an ionic bond compared to a covalent bond?

5. How do cations differ from anions?

6. What chemical feature is used in the classification of minerals into groups?

7. Name the mineral group for the following minerals:

a. Calcite	e. Biotite	i. Pyrite
b. Gypsum	f. Galena	j. Orthoclase
c. Hematite	g. Graphite	k. Magnetite
d. Quartz	h. Fluorite	l. Olivine

8. What is the net charge on an unbonded silica tetrahedron?
9. What allows magnesium to substitute freely for iron in olivine?
10. How are the silica tetrahedra structured differently in pyroxene and amphibole?
11. What are the names and compositions of the two end-members of the plagioclase series?
12. Why does quartz have no additional cations (other than Si^{+4})?
13. Why is colour not necessarily a useful guide to mineral identification?
14. You have an unknown mineral that can scratch glass but cannot scratch a porcelain streak plate. What is its approximate hardness?

Answers at the end of the chapter

References

Klein, C. & Hurlbut, C. S., Jr. (1993). *Manual of Mineralogy (after J. D. Dana).* New York, NY: John Wiley & Sons, Inc.

Nickel, E. H. (1995). The Definition of a Mineral. *The Canadian Mineralogist 33*, 698-690. http://www.minsocam.org/msa/IMA/ima98(04).pdf

Williams, P. (2010, July 28). *Deadliest place on Earth? Surviving Cueva de los Cristales - The Giant Crystal Cave.* http://www.ironammonite.com/2009/12/surviving-cueva-de-los-cristales-giant.html

Answers to Review Questions

1. Charges: proton: +1, neutron: 0, electron: -1, Masses: proton: 1, neutron: 1, electron: almost 0.
2. The element's atomic number will determine the extent to which its outer layers are populated with electrons. If the outer shell is not quite full, the atom may gain electrons to fill them and become an anion (negative charge). If the outer shell has only a few electrons, it may lose them and become a cation (positive charge). Cations and anions attract each other to form molecules with ionic bonding.
3. Helium and neon (and the other noble gases) have complete outer shells and therefore no tendency to form ionic bonds.
4. Electrons are transferred from one atom to another to form an ionic bond. Electrons are shared between atoms to form a covalent bond.
5. An anion has a negative charge and a cation has a positive charge.
6. Minerals are classified into groups based on their anion or anion group.
7. Name the mineral group for the following minerals:

a. Calcite: carbonate	e. Biotite: silicate	i. Pyrite: sulphide
b. Gypsum: sulphate	f. Galena: sulphide	j. Orthoclase: silicate
c. Hematite: oxide	g. Graphite: native min.	k. Magnetite: oxide
d. Quartz: silicate	h. Fluorite: halide	l. Olivine: silicate

8. An unbonded silica tetrahedron has one Si ion (+4 charge) and 4 oxygens (-2 charge each) so the overall charge is 4 – 8 = -4 for SiO_4^{-4}
9. Magnesium can substitute freely for iron in olivine and several other minerals because they have similar charges (+2) and similar ionic radii.
10. Pyroxene is made up of single chains of tetrahedra while amphibole is made up of double chains.
11. The two end-members of the plagioclase series are Albite ($NaAlSi_3O_8$) and Anorthite ($CaAl_2Si_2O_8$)
12. In quartz each silica tetrahedron is bonded to four other tetrahedra, and because oxygens are shared at each bond the overall ratio is silicon (+4) to two oxygens (2 x -2 = -4), which is balanced.
13. Some minerals have distinctive colours, but many have a wide range of colours due to differing impurities.
14. Glass has a Mohs hardness of about 5.5 while porcelain is close to 6.5. The mineral is between these two, so it must be close to 6.

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-NC-SA 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 6. The Rock Cycle

Adapted by Karla Panchuk from *Physical Geology* by Steven Earle

Figure 6.1 A petrified beach near Rock Springs, Wisconsin, U. S. A. The wrinkled face of this vertical cliff displays ripples from an ancient beach. Flowing water moved sand grains to form ripples, and over time the sand was transformed into a solid sedimentary rock. The petrified beach was buried deeper and deeper, and the higher pressures and temperatures caused the sand grains to lose their individual boundaries and merge together. Thus, the sedimentary rock was transformed into a different type of rock, called a metamorphic rock. *Source: Karla Panchuk (2017) CC BY-SA 4.0. See Appendix C for more attributions.*

Learning Objectives

After reading this chapter and answering the *Review Questions* at the end, you should be able to:

- Explain what a rock is.
- Summarize the main characteristics of igneous, sedimentary, and metamorphic rocks.
- Describe the rock cycle and the types of processes that lead to the formation of igneous, sedimentary, and metamorphic rocks.
- Explain why there is an active rock cycle on Earth.

6.1 What Is A Rock?

A **rock** is a solid mass of geological materials. Geological materials include individual mineral crystals, inorganic non-mineral solids like glass, pieces broken from other rocks, and even fossils. The geological materials in rocks may be inorganic, but they can also include organic materials such as the partially decomposed plant matter preserved in coal. A rock can be composed of only one type of geological

material or mineral, but many are composed of several types. Figure 6.1 shows a rock made of three different kinds of minerals.

Figure 6.2 This close-up view of the igneous rock pegmatite shows black biotite crystals, colourless quartz crystals, and pink potassium feldspar crystals. Crystals are mm to cm in scale. *Source: R. Weller/ Cochise College (2011) Permission for non-commercial educational use. (labels added) See Appendix C for more attributions.*

Rocks are grouped into three main categories based on how they form. **Igneous rocks** form when melted rock cools and solidifies. **Sedimentary rocks** form when fragments of other rocks are buried, compressed, and cemented together; or when minerals precipitate from solution, either directly or with the help of an organism. **Metamorphic rocks** form when heat and pressure alter a pre-existing rock. Although temperatures can be very high, metamorphism does not involve melting of the rock.

6.2 The Rock Cycle

The rock components of the crust are slowly but constantly being changed from one form to another. The processes involved are summarized in the **rock cycle** (Figure 6.3). The rock cycle is driven by two forces:

1. Earth's internal heat, which causes material to move around in the core and mantle, driving plate tectonics.
2. The **hydrological cycle**– movement of water, ice, and air at the surface. The hydrological cycle is powered by the sun.

The rock cycle is still active on Earth because our core is hot enough to keep the mantle moving, the atmosphere is relatively thick, and there is liquid water. On some other planets or their satellites (e.g., Mercury), the rock cycle is virtually dead because the core is no longer hot enough to drive mantle convection, and there is no atmosphere or liquid water.

We can start anywhere we like to describe the rock cycle, but it's convenient to start with magma. **Magma** is melted rock located within the Earth. Rock can melt at between about 800 °C and 1300 °C, depending on the minerals in the rock, and the pressure the rock is under. If it cools slowly within the Earth (over centuries to millions of years), magma forms **intrusive igneous rocks**. If magma erupts onto the surface, we refer to it as **lava**. Lava cools rapidly on Earth's surface (within seconds to years) and forms **extrusive igneous rocks** (Figure 6.4).

Figure 6.3 The rock cycle describes processes that form the three types of rock: igneous, sedimentary, and metamorphic. These same processes can turn one type of rock into another. *Source: Karla Panchuk (2017) CC BY-SA 4.0. See Appendix C for more attributions.*

Figure 6.4 Lava flowing from Kīlauea Volcano, Hawai`i. *Source: J. D. Griggs, U. S. Geological Survey (1985) Public Domain*

Mountain building lifts rocks upward where they are acted upon by weathering. Weathering includes chemical processes that break rocks apart, as well as physical processes. Figure 6.5 shows the result of rocks in mountains being broken apart when water gets into cracks, freezes, and forces the cracks wider.

Uplift through mountain building is how rocks once buried deep within Earth can be exposed at Earth's surface.

Figure 6.5 Mountains being broken apart by the wedging action of ice near La Madaleta Glacier, Spain. *Source: Luis Paquito (2006) CC BY-SA 2.0*

The weathering products — mostly small rock and mineral fragments — are eroded, transported, and then deposited as **sediments**. Transportation and deposition occur through the action of glaciers, streams, waves, wind, and other agents. Figure 6.6 shows transportation of fine-grained sediment particles by wind during the Great Depression in the 1930s.

Figure 6.6 Wind transports sediment in a dust storm near Okotoks, Alberta, Canada in July of 1933. *Source: Glenbow Museum Archives, File Number NA-2199-1 (1933) Public Domain*

Sediments are deposited in stream channels, lakes, deserts, and the ocean. Some depositional settings result in characteristic sedimentary structures, such as the ripples that formed when flowing water moved sand along the bottom of the South Saskatchewan River (Figure 6.7).

Unless they are re-eroded and moved along, sediments will eventually be buried by more sediments. At depths of hundreds of metres or more, the sediments become compressed, forcing particles closer together. Mineral crystals grow around and between the particles, binding them together (cementing them). The hardened cemented sediments are **sedimentary rock**. Figure 6.8 shows an example of an ancient sedimentary rock in which ripple structures are preserved, and visible in cross-section as wavy lines.

Figure 6.7 Sand ripples along the South Saskatchewan River, near Saskatoon SK (dog for scale.) *Source: Karla Panchuk (2008) CC BY-SA 4.0*

Figure 6.8 Ripples preserved in 1.2 Ga old sandstone. Notice the wavy lines above the coin. This is a side view of the ripples. *Source: Anne Burgess (2008) CC BY-SA 2.0*

Rocks that are buried very deeply within the crust can reach pressures and temperatures much higher than those at which sedimentary rocks form. Existing rocks that are heated up and squeezed under those extreme conditions are transformed into **metamorphic rocks** (Figure 6.9). The transformation to a metamorphic rock can happen through physical changes, such as when the minerals making up an existing rock re-form into larger crystals of the same mineral. It can also happen through chemical changes, when minerals within the rock react to form new minerals.

Figure 6.9 Limestone, a sedimentary rock formed in marine waters, has been altered by metamorphism into this marble visible on Quadra Island, BC. *Source: Steven Earle (2015) CC BY 4.0*

Chapter 1 Summary

The topics covered in this chapter can be summarized as follows:

6.1 What Is a Rock?

A rock is a solid mass of geological materials. Geological materials include individual mineral crystals, inorganic non-mineral solids like glass, pieces broken from other rocks, and even fossils.

6.2 The Rock Cycle

There are three main types of rock. Igneous rocks form when melted rock cools and solidifies. Sedimentary rock forms from fragments of other rocks, or when crystals precipitate from solution. Metamorphic rocks form when existing rocks are altered by heat, pressure, and/or chemical reactions. The rock cycle summarizes the processes that contribute to transformation of rock from one type to another. The rock cycle is driven by Earth's internal heat, and by processes happening at the surface that are driven by solar energy.

Review Questions

1. What processes must take place to transform rocks into sediment?
2. What processes normally take place in the transformation of sediments to sedimentary rock?
3. What are the processes that lead to the formation of a metamorphic rock?

Answers at the end of the chapter

Answers to Review Questions

1. The rock must be exposed at surface. This means uplift and removal of overlying rocks and sediments is required. Once exposed, chemical and/or physical weathering can reduce the rock to smaller loose fragments (sediments). The sediments can be eroded and then transported by a variety of mechanisms.
2. Sediments are buried beneath other sediments, where pressure compacts the sediments, forcing grains closer together. Mineral cement forms around the grains, binding them to each other and into solid rock.
3. Rock is buried deeply in the crust and exposed to very high temperatures and pressures. Under those conditions, a new type of rock is formed when minerals undergo physical changes and chemical reactions.

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-NC-SA 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 7. Igneous Rocks

Adapted by Karla Panchuk from *Physical Geology* by Steven Earle

Figure 7.1 Lava lake of Mount Nyiragongo, a volcano in the Democratic Republic of Congo. Igneous rocks form when melted rock freezes. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photo by Baron Reznik (2015) CC BY-NC-SA 2.0. Click the image for more attributions. See Appendix C for more attributions.*

Learning Objectives

After reading this chapter and answering the review questions at the end, you should be able to:

- Explain partial melting and the geological processes that lead to melting.
- Describe the range of chemical compositions of magmas.
- Discuss the processes that take place during magma cooling, and the order of crystallization in Bowen's reaction series.
- Explain how fractional crystallization and partial melting alter magma composition.
- Classify igneous rocks according to the proportions of minerals within them.
- Describe the origins of aphanitic, phaneritic, and porphyritic textures
- Classify plutons according to their shapes and relationships to surrounding rocks.
- Explain how chilled margins form.

7.1 Magma and How It Forms

Igneous rocks form when melted rock cools. Melted rock originates within Earth as **magma**. Magma compositions vary, but will have eight main elements in different proportions. The most abundant elements are oxygen and silicon, followed by aluminum, iron, calcium, sodium, magnesium, and potassium. These eight elements are also the most abundant in Earth's crust (Figure 7.2). All magmas have varying proportions of lighter elements such as hydrogen, carbon, and sulphur. Lighter elements are converted into gases like water vapour, carbon dioxide, hydrogen sulphide, and sulphur dioxide as the magma cools.

Figure 7.2 Average composition of Earth's crust by mass. *Source: Steven Earle (2016) CC BY 4.0*

Magma composition depends on the composition of the rocks that melted to form the magma, and on the conditions under which the melting happened. Most igneous rock in Earth's crust comes from magmas that formed through **partial melting** of existing rock, either in the upper mantle or the crust. During partial melting, only some of the minerals within a rock melt. This happens because different minerals have different melting temperatures. The melt is less dense than the surrounding rock, and will percolate upward without the source rock having melted completely. The result is magma with a different composition than the original rock. Partial melting produces melt that has more silica than the original rock, because minerals higher in silica have lower melting points.

To see how partial melting works, consider the mix of materials in Figure 7.3a. It contains white blocks of candle wax, black plastic pipe, green beach glass, and pieces of aluminum wire. When the mixture is heated to 50 °C in a warm oven, the wax melts into a clear liquid (Figure 7.3b), but the other materials remain solid. This is partial melting.

When the mixture is heated to 120 °C, the plastic melts and mixes with the wax, but the aluminum and glass still remain solid (Figure 7.3c). This is still considered partial melting because solid materials remain. When the plastic and wax mixture is poured into a separate container and allowed to cool, the resulting solid has a very different composition from the original mixture (Figure 7.3d). The plastic and wax are analogous to more silica-rich minerals with relatively lower melting points than other minerals in the same rock.

Of course, partial melting in the real world isn't as simple as the example in Figure 7.3. Many rocks are much more complex than the four-component system used here. Some mineral components of rocks may have similar melting temperatures, and begin to melt at the same time. The melting temperature of a mineral may change in the presence of other minerals. Also, when rocks melt, the process can take millions of years, unlike the 90 minutes required to melt the pipe and wax in the experiment in Figure 7.3.

Figure 7.3 An experiment to illustrate partial melting. (a) The original components are white candle wax, black plastic pipe, green beach glass, and aluminum wire. (b) After heating to 50˚C for 30 minutes only the wax has melted. (c) After heating to 120˚C for 60 minutes much of the plastic has melted and the two liquids have mixed. (d) The liquid has been poured off and allowed to cool, making a solid with a different overall composition from the original mixture. *Source: Steven Earle (2015) CC BY 4.0*

Why Rocks Melt

The magma that is produced by partial melting is less dense than the surrounding rock. Magma from partial melting of mantle rocks rises upward through the mantle, and may pool at the base of the crust, or rise through the crust. Moving magma carries heat with it, and some of that heat is transferred to surrounding rocks. If the melting temperature of a rock is less than the temperature of the magma, the rock will begin to melt, and the composition of the magma may change to reflect a mixture of sources. But adding heat is not the only way to trigger melting.

Decompression Melting

Earth's mantle is almost entirely solid rock, in spite of temperatures that would cause rock at Earth's surface to melt. Mantle rock remains solid at those temperatures because the rock is under high pressure. This means that melting can be triggered *without* adding heat if the rock is already hot enough, and the pressure is reduced (Figure 7.4, left, white dashed boxes). Melting triggered by a reduction in pressure is called **decompression melting**.

Figure 7.4 Melting triggers. Left- Decompression melting occurs when rock rises or the overlying crust thins. Right- Flux-induced melting occurs when volatile compounds such as water are added. *Source: Karla Panchuk (2018) CC BY 4.0. Modified after Steven Earle (2016) CC BY 4.0 CC BY 4.0.*

Pressure is reduced when mantle rocks move upward due to convection, or rise as a plume within the mantle. Pressure is also reduced where the crust thins, such as along rift zones.

Flux-induced Melting

When a substance such as water is added to hot rocks, the melting points of the minerals within those rocks decreases. If a rock is already close to its melting point, the effect of adding water can be enough to trigger partial melting. The added water is a **flux**, and this type of melting is called **flux-induced melting**. In Figure 7.4 (right), the rock (represented by the dashed box) is not hot enough to be right of the line where dry mantle rocks melt, but it is to the right of the line where wet mantle rocks melt.

Flux-induced partial melting of rock happens in subduction zones. Minerals are transformed by chemical reactions under high pressures and temperatures, and a by-product of those transformations is water. Relatively little water is required to trigger partial melting. In laboratory studies of the conditions of partial melting in the Japanese volcanic arc, rocks with only 0.2% of their weight consisting of water melted by up to 25%.

Cooling Magma Becomes More Viscous

Viscosity refers to the ease with which a substance flows. A substance with low viscosity is runnier than a substance with high viscosity. At temperatures over 1300°C, most magma is entirely liquid because there is too much energy for the atoms to bond together. As magma loses heat to the surrounding rocks and its temperature drops, things start to change. Silicon and oxygen combine to form silica tetrahedra. With further cooling, the tetrahedra start to link together into chains, or **polymerize**. These silica chains make the magma more viscous. Magma viscosity has important implications for the characteristics of volcanic eruptions.

Exercise: Making Magma Viscous

This is a quick and easy experiment that you can do at home to help you understand the properties of magma. It will only take about 15 minutes, and all you need is half a cup of water and a few tablespoons of flour.

If you've ever made gravy, white sauce, or roux, you'll know how this works.

Figure 7.5 Thick mixture of flour and water. *Source: Steven Earle (2016) CC BY 4.0*

Place 1/2 cup (125 mL) of water in a saucepan over medium heat. Add 2 teaspoons (10 mL) of white flour and stir while continuing to heat the mixture until boiling. The white flour represents silica. The mixture should thicken like gravy because the gluten in the flour becomes polymerized into chains during this process.

Now add more "silica" to see how this changes the viscosity of your magma: take another 4 teaspoons (20 mL) of flour and mix it thoroughly with 4 teaspoons (20 mL) of water in a cup. Add that mixture to the rest of the water and flour in the saucepan. Stir while bringing it back up to nearly boiling temperature, and then allow it to cool. This mixture should slowly become much thicker (Figure 7.5) because there is more gluten, and more chains have formed.

7.2 Crystallization of Magma

The minerals that make up igneous rocks crystallize (solidify, freeze) at a range of different temperatures. This explains why cooling magma can have some crystals within it and yet remain predominantly liquid. The sequence in which minerals crystallize from a magma as it cools is known as **Bowen's reaction series** (Figure 7.6).

Figure 7.6 Bowen's reaction series describe the sequence in which minerals form as magma cools. *Source: Steven Earle (2016) CC BY 4.0.*

How Did We Get Bowen's Reaction Series?

Understanding how the reaction series was derived is key to understanding what it means.

Norman Levi Bowen (Figure 7.7) was born in Kingston Ontario. He studied geology at Queen's University and then at Massachusetts Institute of Technology. In 1912 he joined the Carnegie Institution in Washington, D.C., where he carried out ground-breaking experiments into how magma cools.

Figure 7.7 Norman Bowen in his laboratory. *Source: University of Chicago Photographic Archive, apf1-00841, Special Collections Research Center, University of Chicago Library.*

Working mostly with mafic magmas (magmas rich in iron and magnesium), he determined the order of crystallization of minerals as the temperature drops. First, he melted the rock completely in a specially made kiln. Then he allowed it to cool slowly to a specific temperature before quenching (cooling it quickly) so that no new minerals could form. The rocks that formed were studied under the microscope and analyzed chemically. This was done over and over, each time allowing the magma to cool to a lower temperature before quenching.

The result of these experiments was the reaction series that, even a century later, is still an important basis for our understanding of igneous rocks.

Discontinuous and Continuous Series

Bowen's reaction series (Figure 7.6) has two pathways for minerals to form as magma cools: on the left is the **discontinuous series**. This refers to the fact that one mineral is transformed into a different mineral through chemical reactions. On the right is the **continuous series**, where plagioclase feldspar goes from being rich in calcium to being rich in sodium.

Continuous Series

At about the point where pyroxene begins to crystallize, plagioclase feldspar also begins to crystallize. At that temperature, the plagioclase is calcium-rich (toward the anorthite end-member). As the temperature drops, and providing that there is sodium left in the magma, the plagioclase that forms is a more sodium-rich variety (toward the albite end-member). The series is continuous because the mineral is always plagioclase feldspar, but the series involves a transition from calcium-rich to sodium-rich.

When cooling happens relatively quickly, instead of getting crystals which are of uniform composition, individual plagioclase crystals can be **zoned** from calcium-rich in the centre to more sodium-rich around the outside (Figure 7.8). This occurs because calcium-rich early-forming plagioclase crystals become coated with progressively more sodium-rich plagioclase as the magma cools.

Figure 7.8 Plagioclase crystal exhibiting compositional zones. *Source: Akademia Górniczo-Hutnicza w Krakowie Otwartych Zasobów Edukacyjnych (n.d.) CC BY-NC-SA*

Discontinuous Series

Olivine begins to form at just below 1300°C, but as the temperature drops, olivine becomes unstable. The early-forming olivine crystals react with silica in the remaining liquid and are converted into pyroxene, something like this:

$$\underset{\text{olivine}}{Mg_2SiO_4} + \underset{\text{silica}}{SiO_2} \rightarrow \underset{\text{pyroxene}}{2MgSiO_3}$$

As long as there is silica remaining and the rate of cooling is slow, this process continues down the discontinuous branch: olivine reacts to form pyroxene, and the pyroxene reacts to form amphibole. Under the right conditions amphibole will form to biotite. Finally, if the magma is quite silica-rich to begin with, there will still be some left at around 750 °C to 800 °C, and from this last magma, potassium feldspar, quartz, and maybe muscovite mica will form.

Notice that the sequence of minerals that form goes from isolated tetrahedra (olivine) toward increasingly complex arrangements of silica tetrahedra. Pyroxene consists of single chains, amphibole has double chains, mica has sheets of tetrahedra, and potassium feldspar and quartz at the bottom of the series have tetrahedra connected to each other in three dimensions.

If the magma cools enough, the first minerals to form will be completely used up in later chemical reactions. This is why igneous rocks do not normally have both olivine (at the top of the series) and quartz (at the bottom). Exceptions can occur when rocks that crystallized early in the series come into contact with magmas representing compositions later in the series, such as with the dark green olivine-rich **xenoliths** included within the quartz- and feldspar-rich rock in Figure 7.9. The dark line around the xenoliths is amphibole, which formed as the olivine reacted with the melt. In some of the smaller xenoliths within this boulder, the olivine has been completely transformed into amphibole.

Figure 7.9 Boulder with olivine-rich xenoliths surrounded by silica-rich rock. Black rims on the xenoliths are where the olivine has reacted with the silica-rich melt, forming amphibole. Right- Enlarged view of the amphibole reaction rim. *Source: Karla Panchuk (2018) CC BY 4.0*

Magma Composition: Mafic, Intermediate, and Felsic

The composition of the original magma determines how far the reaction process can continue before all of the magma is used up. In other words, it determines which minerals will form. Magma compositions are reported in terms of the fraction of mass of oxides (e.g., Al_2O_3 rather than just Al; Figure 7.10). On average, **mafic**[1] magma (Figure 7.10, left) is approximately half SiO_2 by mass, and more than 25% iron, magnesium, and calcium oxides by mass. Average **felsic**[2] magmas (Figure 7.10, right) are closer to 75% SiO_2 by mass, and have approximately 5% iron, magnesium, and calcium oxides. Sodium and potassium oxides account for approximately 10% of felsic magmas by mass, but only 5% of mafic magmas. Magmas that fall between mafic and felsic magmas have an **intermediate** composition (Figure 7.10, centre).

Figure 7.10 Chemical compositions of typical mafic, intermediate, and felsic magmas.
Source: Karla Panchuk (2018) CC BY 4.0 modified after Steven Earle (2016) CC BY 4.0.

What Determines the Composition of Magma?

Why Is There No Ultramafic Magma Anymore?

Refer back to Bowen's reaction series in Figure 7.6. Notice that on the far right-hand side of the diagram under "Rock Types," mafic, intermediate, and felsic magma compositions are listed. At the very top of the list is **ultramafic**. Ultramafic rocks have higher MgO than mafic rocks, and even less SiO_2.

The vast majority of silicate rocks in Earth's lithosphere are ultramafic rocks, because the mantle is composed of ultramafic rock. However, ultramafic magma is not encountered in modern volcanic environments, and ultramafic rocks are relatively rare at Earth's surface. The reason is that although Earth was once hot enough to have ultramafic magma, it is no longer hot enough to melt ultramafic rocks. Ultramafic volcanic rocks—called *komatiites*—do exist, but with two notable exceptions, the youngest of these is 2 billion years old.[3]

Oxide	Felsic Magma	Intermediate Magma	Mafic Magma
SiO_2	65-75%	55-65%	45-55%
Al_2O_3	12-16%	14-18%	14-18%
FeO	2-4%	4-8%	8-12%
CaO	1-4%	4-7%	7-11%
MgO	0-3%	2-6%	5-9%
Na_2O	2-6%	3-7%	1-3%

[1] "Mafic" combines the words **MA**gnesium and **F**err**IC** (containing iron).
[2] "Felsic" combines the words **FEL**dspar and **SI**li**C**a.
[3] The komatiites of the Song Da zone in northwestern Vietnam are 270 million years old, and those on Gorgona Island, Columbia are 89 million years old. Exactly how they formed is still a bit of a mystery. See Table 1 of arXiv:physics/0512118v2 [physics.geo-ph] for a compilation of komatiite ages with references.

K_2O	3-5%	2-4%	0.5-3%

Exercise: Mafic, Intermediate, or Felsic?

The proportions of the main chemical components of felsic, intermediate, and mafic magmas are listed in the table at the right. (The values are similar to those shown in Figure 7.10).

Chemical composition by mass for four rock samples are shown in the table below. Compare these with those in the table at the right to determine whether each of these samples is felsic, intermediate, or mafic.

SiO_2	Al_2O_3	FeO	CaO	MgO	Na_2O	K_2O	Type?
55%	17%	5%	6%	3%	4%	3%	
74%	14%	3%	3%	0.5%	5%	4%	
47%	14%	8%	10%	8%	1%	2%	
65%	14%	4%	5%	4%	3%	3%	

Partial Melting Makes Magma That Is Richer in Silica

In partial melting, some components of a mixture melt before others do. In the case of mafic magma, it is produced when ultramafic rocks undergo partial melting. In general, silicate minerals with more silica will melt before those with less silica. This means the partial melt will have more silica than the rock as a whole.

Fractional Crystallization Also Makes Magma Richer In Silica

A number of processes that take place within a magma chamber can affect the types of rocks that form once magma cools and crystallizes. If the magma has a low viscosity— which is likely if the magma is mafic— the crystals that form early, such as olivine (Figure 7.11a), may slowly settle toward the bottom of the magma chamber (Figure 7.11b). This process is called **fractional crystallization**.

The formation of olivine removes iron- and magnesium-rich components, leaving the overall composition of the magma near the top of the magma chamber more felsic. The crystals that settle might either form an olivine-rich layer near the bottom of the magma chamber. Or, because the lower part of the magma chamber is likely to be hotter than the upper part, the crystals might remelt. If remelting happens, crystal settling will make the magma at the bottom of the chamber more mafic than it was to begin with (Figure 7.11c).

Figure 7.11 Formation of a zoned magma chamber. a- Olivine crystals form. b- Olivine crystals settle to the base of the magma chamber, leaving the upper part of the chamber richer in silica. c- Olivine crystals remelt, making magma at the base of the chamber more mafic. *Source: Steven Earle (2015) CC BY 4.0.*

Magma Composition Also Changes When Other Rocks Are Melted And Mixed In

Magma chambers aren't isolated from their surroundings. If the rock in which the magma chamber is located (called the **country rock**) is more felsic than the magma, the country rock may also melt, adding to the magma already in the magma chamber (Figure 7.12). Sometimes magma carries fragments of unmelted rock, called **xenoliths**, within it. Melting of xenoliths can also alter the composition of magma, as can re-melting of crystals that have settled out of the magma.

Figure 7.12 The composition of magma in a magma chamber is affected by fractional crystallization within the magma chamber, but it can also be affected by partial melting of the rock surrounding the magma chamber, melting of xenoliths within the magma, or re-melting of crystals that have settled to the bottom of the magma chamber. *Source: Steven Earle (2015) CC BY 4.0*

7.3 Classification of Igneous Rocks

Classification By Mineral Abundance

Igneous rocks can be divided into four categories based on their chemical composition: felsic, intermediate, mafic, and ultramafic. The diagram of Bowen's reaction series (Figure 7.6) shows that differences in chemical composition correspond to differences in the types of minerals within an igneous rock. Igneous rocks are given names based on the proportion of different minerals they contain. Figure 7.13 is a diagram with the minerals from Bowen's reaction series, and is used to decide which name to give an igneous rock.

Figure 7.13 Classification diagram for igneous rocks. Igneous rocks are classified according to the relative abundances of minerals they contain. A given rock is represented by a vertical line in the diagram. In the ultramafic field, the arrows represent a rock containing 48% pyroxene and 52% plagioclase feldspar. The name an igneous rock gets depends not only on composition, but on whether it is intrusive or extrusive. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0, modified after Steven Earle (2015) CC BY 4.0 and others, with photos by R. Weller/Cochise College. See Appendix C for links to photos and notes on diagram construction.*

To see how Figure 7.13 works, first notice the scale in percent along the vertical axis. The interval between each tick mark represents 10% of the minerals within a rock by volume. An igneous rock can be represented as a vertical line drawn through the diagram, and the vertical scale used to break down the proportion of each mineral it contains. For example, the arrows in the ultramafic field of the diagram represent a rock containing 48% pyroxene and 52% plagioclase feldspar. An igneous rock at the boundary between the mafic and ultramafic fields (marked with a vertical dashed line) would have approximately 20% olivine, 50% pyroxene, and 30% Ca-rich plagioclase feldspar by volume.

Classification By Grain Size

The name an igneous rock gets also depends on whether it cools within Earth (an **intrusive** or **plutonic** igneous rock), or whether it cools on the Earth's surface after erupting from a volcano (an **extrusive** or **volcanic** igneous rock). For example, a felsic *intrusive* rock is called **granite**, whereas a felsic *extrusive* rock is called **rhyolite**. Granite and rhyolite have the same mineral composition, but their grain size gives each a distinct appearance.

The key difference between intrusive and extrusive igneous rocks—the size of crystals making them up—is related to how rapidly melted rock cools. The longer melted rock has to cool, the larger the crystals within it can become. Magma cools much slower within Earth than on Earth's surface because magma within Earth is insulated by surrounding rock. Notice that in Figure 7.13, the intrusive rocks have crystals large enough that you can see individual crystals—either by identifying their boundaries, or seeing light reflecting from a crystal face. A rock with individual crystals that are visible to the unaided eye has a **phaneritic** or **coarse-grained** texture. The extrusive rocks in the second row have much smaller crystals. The crystals are so small that individual crystals cannot be distinguished, and the rock looks like a dull mass. A rock with crystals that are too small to see with the unaided eye has an **aphanitic** or **fine-grained** texture. Table 7.1 summarizes the key differences between intrusive and extrusive igneous rocks.

Table 7.1 Comparison of intrusive and extrusive igneous rocks

	Magma cools within the Earth	**Lava cools on Earth's surface**
Terminology	Intrusive/ plutonic	Extrusive/ volcanic
Cooling rate	Slow: surrounding rocks insulate the magma chamber	Rapid: heat is exchanged with the atmosphere
Texture	**Phaneritic** (coarse-grained): individual crystals are large enough to see without magnification.	**Aphanitic** (fine-grained): crystals are too small to see without significant magnification (fine-grained)

What this means is that two igneous rocks comprised of exactly the same minerals, and in the exactly the same proportions, can have different names. A rock of intermediate composition is **diorite** if it is course-grained, and **andesite** if it is fine-grained. A mafic rock is **gabbro** if it is course-grained, and **basalt** if fine-grained. The course-grained version of an ultramafic rock is **peridotite**, and the fine-grained version is **komatiite**. It makes sense to use different names because rocks of different grain sizes form in different ways and in different geological settings.

Does This Mean an Igneous Rock Can Only Have One Grain Size?

No. Something interesting happens when there is a change in the rate at which melted rock is cooling. If magma is cooling in a magma chamber, some minerals will begin to crystallize before others do. If cooling is slow enough, those crystals can become quite large.

Now imagine the magma is suddenly heaved out of the magma chamber and erupted from a volcano. The larger crystals will flow out with the lava. The lava will then cool rapidly, and the larger crystals will be surrounded by much smaller ones. An igneous rock with crystals of distinctly different size (Figure 7.14) is said to have a **porphyritic** texture, or might be referred to as a **porphyry**. The larger crystals are called **phenocrysts**, and the smaller ones are referred to as the **groundmass**.

Figure 7.14 Porphyritic rhyolite with quartz and potassium feldspar phenocrysts within a dark groundmass. Porphyritic texture (when different crystal sizes are present) is an indication that melted rock did not cool at a constant rate. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photo by R. Weller/Cochise College (2011)*

Exercise: Which Mineral Will the Phenocryst Be?

As a magma cools below 1300°C, minerals start to crystallize within it. If the magma is then erupted, the rest of the liquid will cool quickly to form a **porphyritic** texture. The rock will have some relatively large crystals (**phenocrysts**) of the minerals that crystallized early, and the rest will be very fine-grained or even glassy. Using the diagram shown here, predict what phenocrysts might be present where the magma cooled as far as line **a**. Which would be present where magma cooled to line **b**?

Figure 7.15 Bowen's reaction series. *Source: Steven Earle (2015) CC BY 4.0*

Classifying Igneous Rocks According to the Proportion of Dark Minerals

If you unsure of which minerals are present in an intrusive igneous rock, there is a quick way to approximate the composition of that rock. In general, igneous rocks have an increasing proportion of dark minerals as they become more mafic (Figure 7.16).

The dark-coloured minerals are those higher in iron and magnesium (e.g., olivine, pyroxene, amphibole, biotite), and for that reason they are sometimes referred to collectively as **ferromagnesian** minerals. By estimating the proportion of light minerals to dark minerals in a sample, it is possible to place that sample in Figure 7.16. Graphical scales are used to help visualize the proportions of light and dark minerals (Figure 7.17).

Figure 7.16 Simplified igneous rock classification according to the proportion of light and dark (or ferromagnesian) minerals. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0.*

Figure 7.17 A guide for estimating the proportion of dark minerals in an igneous rock. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0*

It is important to note that estimating the proportion of dark minerals is only approximate as a means for identifying igneous rocks. One problem is that plagioclase feldspar is light-coloured when it is sodium-rich, but can appear darker if it is calcium-rich. Plagioclase feldspar is not ferromagnesian, so it falls in the non-ferromagnesian (light minerals) region in Figure 7.16 even when it has a darker colour.

Exercise: Classifying Igneous Rocks by the Proportion of Dark Minerals

The four igneous rocks shown below have differing proportions of ferromagnesian silicates (dark minerals). Estimate the proportion of dark minerals using the guide in Figure 7.17, and then use Figure 7.16 to determine the likely rock name for each one.

Figure 7.18 Identify these rocks by estimating the proportion of dark minerals in each sample. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0*

Classifying Igneous Rocks When Individual Crystals Are Not Visible

The method of estimating the percentage of minerals works well for phaneritic igneous rocks, in which individual crystals are visible with little to no magnification. If an igneous rock is porphyritic but otherwise aphanitic (e.g., Figure 7.14), the minerals present as phenocrysts give clues to the identity of the rock. However, there are cases where mineral composition cannot be determined by looking at visible crystals. These include volcanic rocks without phenocrysts, and glassy igneous rocks.

Volcanic Rocks Without Phenocrysts

In the absence of visible crystals or phenocrysts, volcanic rocks are be classified on the basis of colour and other textural features. As you may have noticed in Figure 7.13, the colour of volcanic rocks goes from light to dark as the composition goes from felsic to mafic. Rhyolite is often a tan or pinkish colour, andesite is often grey, and basalt ranges from brown to dark green to black (Figure 7.19).

Figure 7.19 In volcanic igneous rocks, individual crystals are not visible. Colours change from light to dark as the composition of the rocks go from felsic to mafic. Vesicles and amygdules are common characteristics of basalt. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photos by R. Weller/ Cochise College. See Appendix C for links to photos.*

Basalt often shows textural features related to lava freezing around gas bubbles. When magma is underground, pressure keeps gases dissolved, but once magma has erupted, the pressure is much lower. Gases dissolved in the lava are released, and bubbles can develop. When lava freezes around the bubbles, **vesicles** are formed (circular inset in 7.19). If the vesicles are later filled by other minerals, the filled vesicles are called **amygdules** (box inset in Figure 7.19).

Glassy Volcanic Rocks

Crystal size is a function of cooling rate. The faster magma or lava cools, the smaller the crystals it contains. It is possible for lava to cool so rapidly that no crystals can form. The result is called **volcanic glass**. Volcanic glass can be smooth like **obsidian** or vesicular like **scoria** (mafic) and **pumice** (felsic; Figure 7.20). Pumice can float on water because of its low-density felsic composition and enclosed vesicles.

Figure 7.20 Glassy volcanic rocks. Obsidian has a glassy lustre, but scoria and pumice are highly vesicular. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photos by R. Weller/Cochise College. See Appendix C for links to photos.*

7.4 Intrusive Igneous Rocks

In most cases, a body of hot magma is less dense than the rock surrounding it, so it has a tendency to creep upward toward the surface. It does so in a few different ways:

- Filling and widening existing cracks
- Melting the surrounding rock
- Pushing the rock aside (where the rock is hot enough and under enough pressure to deform without breaking)
- Breaking the rock

Figure 7.21 Xenoliths of mafic rock in granite, Victoria, B.C. The fragments of dark rock have been broken off and incorporated into the light-coloured granite. *Source: Steven Earle (2015) CC BY 4.0*

When magma forces itself into cracks, breaks off pieces of rock, and then envelops them, this is called **stoping**. The resulting fragments are xenoliths.[4] Xenoliths may appear as dark patches within a rock (Figure 7.21).

Some of the magma may reach the surface, resulting in volcanic eruptions, but most cools within the crust. The resulting body of rock is called a **pluton**.[5] Plutons can have different shapes and different relationships with the surrounding country rock (Figure 7.22). These characteristics determine what name the pluton is given.

Large, irregularly shaped plutons are called **stocks** or **batholiths**, depending upon their size. Tabular plutons are called **dikes** if they cut across existing structures, and **sills** if they are parallel to existing structures. **Laccoliths** are like sills, except they have caused the overlying rocks to bulge upward. **Pipes** are cylindrical conduits.

Figure 7.22 Plutons can have a variety of shapes, and be positioned in a variety of ways relative to the surrounding rocks. They are named according to these characteristics. *Source: Karla Panchuk (2018) CC BY 4.0*

---

[4] From the Greek words *xenos*, meaning "foreigner" or "stranger," and *lithos* for "stone."

[5] After Pluto was demoted from planet status, astronomers tried to come up with a name for objects like Pluto. For a while they considered "pluton" however geologists rightly objected that they had first claim on the word. In the end the International Astronomical Union settled on "dwarf planet" instead.

Types of Plutons

Stocks and Batholiths

Large irregular-shaped plutons are called either **stocks** or **batholiths**, depending on their area. If an irregularly shaped body has an area greater than 100 km^2, then it's a batholith, otherwise it's a stock. Note that our knowledge of the size of a body can be limited to what we see at the surface. A body with an area of less than 100 km^2 exposed at the surface might in fact be much larger at depth. It might be classified as a stock initially, until someone is able to map out its true extent.

Batholiths are typically formed when a number of stocks coalesce beneath the surface to create one large body. One of the largest batholiths in the world is the Coast Range Plutonic Complex[6], which extends all the way from the Vancouver region to southeastern Alaska (Figure 7.23).

Tabular Intrusions

Tabular (sheet-like) plutons are classified according to whether or not they are **concordant** with (parallel to) existing layering (e.g., sedimentary bedding or metamorphic foliation[7] in the country rock. A **sill** is concordant with existing layering, and a **dike**[8] is **discordant**. If the country rock has no bedding or foliation, then any tabular body within it is a dike. Note that the sill-versus-dike designation is not determined simply by the orientation of the feature. A dike could be horizontal and a sill could be vertical- it all depends on the orientation of features in the surrounding rocks.

A **laccolith** is a sill-like body that has expanded upward by deforming the overlying rock. If a sill forms, but magma pools and sags downward, it creates a **lopolith**.

Figure 7.23 The Coast Range Plutonic Complex (also called the Coast Range Batholith) is the largest in the world. It is part of a chain of batholiths along the western coast of North America. *Source: Karla Panchuk (2018) CC BY 4.0. Modified after Bally (1989).*

Pipes

A **pipe**, as the name suggests, is a cylindrical body with a circular, elliptical, or even irregular cross-section, that serves as a conduit (or pipeline) for the movement of magma from one location to another. Pipes may feed volcanoes, but pipes can also connect plutons.

Chilled Margins

As discussed already, plutons can interact with the rocks into which they are intruded. Partial melting of the country rock may occur, or stoping may form xenoliths. The heat from magma can even cause causing mineralogical and textural changes in country rock. However, country rock can also affect the magma.

The most obvious effect that country rock can have on magma is a **chilled margin** along the edges of the pluton (Figure 7.24). The country rock is much cooler than the magma, so magma that comes into contact with the country rock cools faster than magma toward the interior of the pluton. Rapid cooling leads to

---

[6] Also referred to as the Coast Range Batholith.
[7] Sedimentary bedding refers to the layers in which sedimentary rocks form. Metamorphic foliation refers to the way minerals or other elements in a rock are aligned as a result of being deformed by heat and pressure. Bedding and foliation will be discussed in more detail in later chapters.
[8] Also spelled *dyke*.

smaller crystals, so the texture along the edges of the pluton is different from that of the interior of the pluton, and the colour may be darker.

Figure 7.24 A mafic dyke with chilled margins within basalt at Nanoose, B.C. The coin is 24 mm in diameter. The dyke is about 25 cm across and the chilled margins are 2 cm wide. *Source: Steven Earle (2015) CC BY 4.0*

Exercise: Pluton Problems

The diagram below is a cross-section through part of the crust showing a variety of intrusive igneous rocks. Indicate whether each of the plutons labelled *a* to *e* on the diagram below is a *dike*, a *sill*, a *stock*, or a *batholith*. (Note the trees for scale.)

Figure 7.15 A variety of igneous intrusions. *Source: Steven Earle (2015) CC BY 4.0*

Chapter 7 Summary

The topics covered in this chapter can be summarized as follows:

7.1 Magma and How It Forms

Magma is molten rock, and in most cases, it forms from partial melting of existing rock. The chemistry of magma depends on the source rock that is melting, as well as the degree of partial melting that occurs. Magma forms by decompression melting, flux-induced melting, and heat transfer. Magmas range in composition from ultramafic to felsic. Mafic rocks are rich in iron, magnesium, and calcium, and contain approximately 50% silica. Felsic rocks are richer in silica (~70%) and have lower levels of iron, magnesium, and calcium, and higher levels of sodium and potassium than mafic rocks.

7.2 Crystallization of Magma

As a body of magma starts to cool, the first process to take place is the polymerization of silica tetrahedra into chains. This increases the magma's viscosity (makes it thicker) and because felsic magmas have more silica than mafic magmas, they tend to be more viscous. Bowen's reaction series allows us to predict the order of crystallization of magma as it cools. Magma can be modified by fractional crystallization (separation of early-forming crystals), by mixing in material from the surrounding rocks by partial melting, and by mixing with magmas of differing chemistry.

7.3 Classification of Igneous Rocks

Igneous rocks are classified based on their mineral composition and texture. Felsic igneous rocks have less than 20% dark minerals (ferromagnesian silicates including amphibole and/or biotite) with varying amounts of quartz, both potassium and plagioclase feldspars, and sometimes muscovite. Mafic igneous rocks have more than 50% dark minerals (primarily pyroxene) plus plagioclase feldspar. Most intrusive igneous rocks are phaneritic (individual crystals are visible unmagnified). If there were two stages of cooling (slow then fast), the texture may be porphyritic (large crystals in a matrix of smaller crystals).

7.4 Intrusive Igneous Bodies

Magma intrudes into country rock by pushing it aside or melting through it. Intrusive igneous bodies tend to be irregular (stocks and batholiths), tabular (dikes and sills), or pipe-like. Batholiths have areas of 100 km^2 and greater, while stocks are smaller. Sills are parallel to existing layering in the country rock, while dikes cut across layering. A pluton that intruded into cold rock is likely to have a chilled margin.

Review Questions

1. What is the significance of the term *reaction* in *Bowen's reaction series*?
2. Why is it common for plagioclase crystals to be zoned from relatively calcium-rich in the middle to more sodium-rich toward the edge?
3. What must happen within a magma chamber for fractional crystallization to take place?
4. Explain the difference between aphanitic and phaneritic textures.
5. Name the following rocks:

    a. An extrusive rock with 40% Ca-rich plagioclase and 60% pyroxene

    b. An intrusive rock with 65% plagioclase, 25% amphibole, and 10% pyroxene

    c. An intrusive rock with 25% quartz, 20% potassium feldspar, 50% plagioclase feldspar, and minor amounts of biotite

6. What is the difference between a concordant tabular intrusion and a discordant tabular intrusion?
7. Why do dikes commonly have fine-grained margins?
8. What is the difference between a batholith and a stock?
9. Describe two ways in which batholiths intrude into existing rock.
10. Why is compositional layering a common feature of mafic plutons but not of felsic plutons?

Answers at the end of the chapter

References

Bally, A. W. (1989). Plate 10. Selected distribution maps, rate of accumulation maps, and lithofacies maps—Phanerozoic, North America. In A. W. Bally & A. R. Palmer (Eds.), *The Geology of North America—An Overview: Volume A*. Boulder: Geological Society of America.

Kushiro, I. (2007). Origins of magmas in subduction zones: a review of experimental studies. *Proceedings of the Japan Academy, Series B, Physical and Biological Sciences 83*(1), 1-15. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3756732/

Answers to Review Questions

1. As the temperature decreases minerals that formed early (e.g., olivine) may react with the remaining magma to form new minerals (e.g., pyroxene).
2. Calcium-rich plagioclase forms early on in the cooling process of a magma, but as the temperature drops, a more sodium-rich variety forms around the existing crystals.
3. Some minerals must begin to form while melt is still present. Early-forming minerals, which are typically quite dense (e.g., olivine), will sink to the bottom of the magma chamber if the magma is not too viscous, thus becoming separated from the rest of the magma. The composition of the remaining magma will be more felsic than before.
4. Phaneritic texture means that individual crystals can be distinguished by the naked eye. Aphanitic texture means that individual crystals cannot be distinguished without a microscope. The dividing line is somewhere between 0.1 and 1 mm, depending on the minerals.
5. a) basalt; b) diorite; c) granite
6. A concordant body (a sill) is parallel to any pre-existing layering (e.g., bedding or foliation) in the country rock is. A discordant body (a dike) cuts across any pre-existing layering, or has intruded in country rock without layering (e.g., in granite).
7. When the hot magma intrudes into cold country rock, the margins cool quickly and small crystals form, whereas magma that is not in contact with the cool country rock will cool more slowly, and larger crystals will form. The chilled margin is the band of small crystals along the edge of the dike.
8. A batholith has an area of 100 km^2 or greater, whereas a stock is smaller.
9. Batholiths (or stocks) intrude into existing rock by (a) melting through the country rock, or (b) causing the country rock to break and fall into the magma (stoping), or (c) pushing the country rock aside.
10. Compositional layering forms when early-crystallizing minerals sink toward the bottom of a magma chamber. This can only happen in non-viscous magma. Mafic magma is typically much less viscous than felsic magma.

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-NC-SA 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 8. Weathering, Sediment, & Soil

Adapted by Karla Panchuk from *Physical Geology* by Steven Earle

Figure 8.1 The Hoodoos, near Drumheller, Alberta, have formed from the differential weathering (weaker rock weathering faster than stronger rock) of sedimentary rock. *Source: Steven Earle (2015) CC BY 4.0.*

Learning Objectives

After reading this chapter and answering the review questions at the end, you should be able to:

- Explain why rocks formed at depth in the crust are susceptible to weathering at the surface.
- Describe the main processes of mechanical weathering, and the materials that are produced.
- Describe the main processes of chemical weathering, and common chemical weathering products.
- Explain the characteristics used to describe sediments, and what those characteristics can tell us about the origins of the sediments.
- Discuss the relationships between weathering and soil formation, and the origins of soil horizons.
- Describe and explain the distribution of Canadian soil types.
- Explain how changing weathering rates affect the carbon cycle and the climate system.

What Is Weathering?

Weathering occurs when rock is exposed to the "weather" — to the forces and conditions that exist at Earth's surface. Rocks that form deep within Earth experience relatively constant temperature, high pressure, have no contact with the atmosphere, and little or no interaction with moving water. Once overlying layers are eroded away and a rock is exposed at the surface, conditions change dramatically. Temperatures vary widely, and pressure is much lower. Reactive gases like oxygen and carbon dioxide are plentiful, and in many climates, water is abundant.

Weathering can be characterized as **mechanical** (or **physical**), and **chemical**. In mechanical weathering, physical processes break rock into smaller pieces. In chemical weathering, chemical reactions change minerals into forms that are less affected by chemical reactions that occur at Earth's surface. Mechanical and chemical weathering reinforce each other, because mechanical weathering provides new fresh surfaces for attack by chemical processes, and chemical weathering weakens the rock so that it is more susceptible to mechanical weathering. Together, these processes create the particles and ions that can eventually become sedimentary rock. They also create soil, which is necessary for our existence on Earth.

8.1 Mechanical Weathering

Intrusive igneous rocks form at depths of 100s of metres to 10s of kilometres. Most metamorphic rocks are formed at depths of kilometres to 10s of kilometres. Sediments are turned into sedimentary rocks only when they are buried by other sediments to depths in excess of several 100s of metres. Weathering cannot happen until these rocks are revealed at Earth's surface by uplift and the erosion of overlying materials. Once the rock is exposed at the surface as an **outcrop**, weathering begins.

The agents of mechanical weathering can be broadly classified into two groups: those that cause the outer layers of a rock to expand, and those that act like wedges to force the rock apart.

Mechanical Weathering By Expansion

Some processes at Earth's surface can cause a thin outer layer of a rock to expand. Deeper than the thin outer layer, the rock does not expand. The difference is accommodated by a crack developing between the outer and inner layers, breaking the outer layer off in slabs (Figures 8.2 and 8.3). When layers break off a rock in slabs or sheets, it is referred to as **exfoliation**.

Figure 8.2 Close-up view of exfoliation of a granite dome in the Enchanted Rock State Natural Area, Texas, USA. *Source: Wing-Chi Poon (2005) CC BY-SA 2.5*

Granite tends to exfoliate parallel to the exposed surface because it does not have planes of weakness to determine how it breaks. In contrast, sedimentary rocks tend to exfoliate along the contacts between different sedimentary layers, and metamorphic rocks tend to exfoliate parallel to aligned minerals.

Figure 8.3 View of exfoliation at a distance (centre of image) in granite exposed on the west side of the Coquihalla Highway north of Hope, B.C. *Source: Steven Earle (2015) CC BY*

Reasons Rocks Expand

A rock within the Earth has pressure exerted upon it by other rocks sitting above it. This is called **confining pressure.** When the overlying mass is removed by weathering, the confining pressure decreases, allowing the rock to expand. The cracking that results is sometimes called **pressure-release cracking**.

Heating a rock can also cause it to expand. If the rock is heated rapidly, as during a wildfire, cracks can form. If it goes through large daily temperature swings (e.g., in the desert where it is very hot during the day but cold at night), cracking can also eventually result as the rock is weakened.

Mechanical Weathering by Wedging

In wedging, a pre-existing crack in a rock is made larger by forcing it open.

Frost Wedging

Frost wedging (or **ice wedging**) happens when water seeps into cracks, then expands upon freezing. The expansion enlarges the cracks (Figure 8.4). The effectiveness of frost wedging depends on how often freezing and thawing occur. Frost wedging won't be as important in warm areas where freezing is infrequent, in very cold areas where thawing is infrequent, or in very dry areas, where there is little water to seep into cracks.

Frost wedging is most effective in Canada's climate, where for at least part of the year temperatures oscillate between warm and freezing. In many parts of Canada, the temperature swings between freezing at night and thawing in the day tens to hundreds of times a year. Even in warm coastal areas of southern British Columbia, freezing and thawing transitions are common at higher elevations. A common feature in areas of effective frost wedging is a talus slope — a fan-shaped deposit of fragments removed by frost wedging from the steep rocky slopes above (Figure 8.5).

Figure 8.4 A rock broken by ice wedging sits in a stream in Mount Revelstoke National Park, Canada. Rocks break apart when ice expands in pre-existing cracks. *Source: Karla Panchuk (2018) CC BY 4.0.*

Figure 8.5 An area with very effective frost wedging near Keremeos, BC. Fragments that were wedged away from the cliffs above have accumulated in a talus deposit at the base of the slope. The rocks in this area are variable in colour, which is reflected in the colours of the talus. *Source: Steven Earle (2015) CC BY 4.0*

Salt Wedging

Salt wedging happens when saltwater seeps into rocks and then evaporates on a hot sunny day. Salt crystals grow within cracks and pores in the rock, and the growth of these crystals can push grains apart, causing the rock to weaken and break. There are many examples of this on the rocky shorelines of Vancouver Island and the Gulf Islands, where sandstone outcrops are common and salty seawater is readily available (Figure 8.6). The honeycomb structure of rounded holes, called **tafoni**, is related to the original roughness of the surface. Low spots collect salt water, causing the effect to be accentuated around existing holes.

Figure 8.6 Tafoni (Honeycomb weathering) in sandstone on Gabriola Island, British Columbia. The holes are caused by crystallization of salt within rock pores. *Source: Steven Earle (2015) CC BY 4.0*

Plant and Animal Activity

The effects of plants are significant in mechanical weathering. Roots can force their way into even the tiniest cracks. They exert tremendous pressure on the rocks as they grow, widening the cracks and breaking the rock. This is called **root wedging** (Figure 8.7).

Figure 8.7 Root wedging along a quarry wall. Left: Rocks beneath the thick red beds have been split into sheets by tree roots. Right: A closer examination reveals that tree roots are working into vertical cracks as well. *Source: Karla Panchuk (2018) CC BY 4.0*

Although most animals do not normally burrow through solid rock, they can excavate and remove huge volumes of soil, and thus expose the rock to weathering by other mechanisms. Humans modify vast tracts of land by excavation, and have a profound effect on accelerating mechanical weathering.

Exercise: Mechanical Weathering

What mechanical weathering processes do you think take place on this mountain in British Columbia?

Figure 8.8 Granite at the top of Siy'ám' Smánit (also known as Stawamus Chief Mountain), near Squamish, British Columbia. *Source: Steven Earle (2015) CC BY 4.0*

Erosion

Mechanical weathering is greatly facilitated by **erosion**. Erosion is the removal of weathering products, such as fragments of rock. This exposes more rock to weathering, accelerating the process. A good example of weathering and erosion working together is the talus shown in Figure 8.5. The rock fragments forming the talus piles were broken off the steep rock faces at the top of the cliff by ice wedging, and then removed by gravity.

Gravity does not always work alone to remove weathering products. Other agents of erosion include water in streams, ice in glaciers, and waves on coasts.

8.2 Chemical Weathering

Chemical weathering results from chemical changes to minerals that become unstable when they are exposed to surface conditions. The kinds of changes that take place are specific to the mineral and the environmental conditions. Some minerals, like quartz, are virtually unaffected by chemical weathering. Others, like feldspar, are easily altered.

Types of Chemical Weathering Reactions

Dissolution

Dissolution reactions produce ions, but no minerals, and are reversible if the solvent is removed. A household example would be dissolving a teaspoon of table salt (the mineral halite) in a glass of water. The halite will separate into Na^+ and Cl^- ions. If the water in the glass is allowed to evaporate, there will not be enough water molecules to hold the Na^+ and Cl^- ions apart, and the ions will come together again to form halite. Gypsum and anhydrite are other minerals that will dissolve in water alone.

Other minerals, such as calcite, will dissolve in acidic water. Acidic water is common in nature, because carbon dioxide (CO_2) in the atmosphere reacts with water vapour in the atmosphere, and with water on land and in the oceans to produce carbonic acid (Figure 8.9).

Figure 8.9 Calcite weathering by dissolution. Top: Carbon dioxide reacts with water to make acid. Bottom: Acid reacts with calcite and produces ions. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Modified after What-When-How. Molecules from JMSE Molecular Editor, Bienfait and Ertl (2013), with permission for CC BY-NC-SA use.*

While rainwater and atmospheric CO_2 can combine to create carbonic acid, the amount of CO_2 in the air is enough to make only very weak carbonic acid. In contrast, biological processes acting in soil can result in a much higher concentration of CO_2 within soil, as well as adding organic acids. Water that percolates through the soil can become significantly more acidic.

Calcite is a major component of the sedimentary rock called limestone (typically more than 95%). In the presence of acidic groundwater, limestone can dissolve underground. Over time the dissolution can remove enough of the calcite to form caves.

If dissolution of limestone or other materials removes enough rock to undermine support near the surface, the surface may collapse, creating a **sinkhole** such as the one in Figure 8.10, downstream of the Mosul Dam in Iraq.

Figure 8.10 Sinkhole downstream of the Mosul Dam in Iraq. The sinkhole is a result of dissolution of gypsum and anhydrite layers. *Source: U. S. Army Corps of Engineers (2007) Public Domain*

Although the sinkhole in Figure 8.9 might appear minor, it indicates a serious problem. The dam itself is constructed on limestone supported by beds of gypsum and anhydrite. Gypsum and anhydrite are soluble in water, and the gypsum and anhydrite beneath the dam are rapidly dissolving away. This was the case prior to construction of the dam. However, once the dam was filled, the increased water pressure began to force water through the formations much faster, accelerating dissolution. Ongoing measures to fill gaps with grout are required, or else there is a grave risk of catastrophic failure, placing nearly 1.5 million people at risk.

Hydrolysis

The term **hydrolysis** combines the prefix *hydro*, referring to water, with *lysis*, which is derived from a Greek word meaning to loosen or dissolve. Thus, you can think of hydrolysis as a chemical reaction where water loosens the chemical bonds within a mineral. This might sound the same as dissolution but the difference is that hydrolysis produces a different mineral in addition to ions. An example of hydrolysis is when water reacts with potassium feldspar to produce **clay minerals** and ions. The results can be seen by comparing weathered and unweathered surfaces of the same sample of granite (Figure 8.11). On the recently broken unweathered surface (Figure 8.11, left) feldspar is visible as bright white crystals. On a weathered surface (right) the feldspar has been altered to the chalky-looking clay mineral kaolinite.

Figure 8.11 A piece of granite with unweathered (left) and weathered (right) surfaces. On the unweathered surfaces the feldspars are still fresh and glassy looking. On the weathered surface there are chalky white patches where feldspar has been altered to the clay mineral kaolinite. *Source: Karla Panchuk (2018) CC BY 4.0. Photos by Steven Earle (2015) CC-BY 4.0*

Silicate minerals other than feldspar can undergo hydrolysis, but with different end results. For example, pyroxene can be converted to the clay minerals chlorite or smectite. Olivine can be converted to the clay mineral serpentine.

Hydration

Hydration reactions involve water being added to the chemical structure of a mineral. An example of a hydration reaction is when anhydrite ($CaSO_4$) is transformed into gypsum ($CaSO_4 \cdot 2H_2O$). A consequence of hydration is that the resulting mineral has a greater volume than the original mineral. In the case of the

Mosul Dam, hydration of anhydrite has important consequences. The increase in volume applied force to an overlying limestone layer, breaking it into pieces. While unbroken limestone is a strong enough material upon which to build a foundation, broken limestone is too weak to provide a safe foundation.

Oxidation

Oxidation happens when free oxygen (i.e., oxygen not bound up in molecules with other elements) is involved in chemical reactions. Oxidation reactions provide valuable insight into Earth's early surface conditions because there is a clear transition in the rock record from rocks containing no minerals that are products of oxidation reactions, to rocks containing abundant minerals produced by oxidation. This reflects a transition from an oxygen-free atmosphere to an oxygenated one.

In iron-rich minerals such as olivine, the oxidation reaction begins with taking iron out of the mineral and putting it into solution as an ion. Olivine reacts with carbonic acid, leaving dissolved iron, bicarbonate, and silicic acid:

$$Fe_2SiO_4 + 4H_2CO_3 \rightarrow 2Fe^{2+} + 4HCO_3^- + H_4SiO_4$$

Iron and oxygen dissolved in water react in the presence of bicarbonate to produce hematite and carbonic acid:

$$2Fe^{2+} + \tfrac{1}{2} O_2 + 2H_2O + 4HCO_3^- \rightarrow Fe_2O_3 + 4H_2CO_3$$

When the olivine in basalt is oxidized, the basalt takes on a reddish colour that is distinct from the dark grey or black of unweathered basalt (Figure 8.12).

Figure 8.12 Basalt pillows in Andalusia, Spain, with reddish weathered surfaces. Where parts of the pillows have broken away, darker unweathered basalt is visible. *Source:Ignacio Benvenuty Cabral (2011) CC BY-NC-SA*

The oxidation reaction would be similar for other iron-containing silicate minerals such as pyroxene, amphibole, and biotite. Iron in sulphide minerals such as pyrite (FeS_2) can also be oxidized in this way.

Hematite is not the only mineral that can result from oxidation. In fact, a wide range of iron oxide minerals that can form in this way, In granite, for example, biotite and amphibole can be altered to form the iron oxide and iron hydroxyoxide minerals that are referred to in combination as **limonite** (orange material in Figure 8.13).

Figure 8.13 Biotite and amphibole in this granite have been altered by oxidation to limonite (orange-yellow coating), which is a mixture of iron oxide and iron hydroxyoxide minerals. *Source: Steven Earle (2015) CC-BY 4.0*

Oxidation Reactions and Acid Rock Drainage

Oxidation reactions can pose an environmental problem in areas where rocks have elevated levels of sulphide minerals such as pyrite. This is because when oxygen and water react with pyrite, sulphuric acid is produced:

$$2FeS_2 + 7O_2 + 2H_2O \rightarrow 2FeSO_4 + 2H_2SO_4$$

The runoff from areas where this process is taking place is known as **acid rock drainage** (ARD), and even a rock with 1% or 2% pyrite can produce significant ARD. Some of the worst examples of ARD are at metal mine sites, especially where pyrite-bearing rock and waste material have been mined from deep underground, and then piled up and left exposed to water and oxygen. In these cases the problem is referred to as **acid mine drainage**. One example is the Mt. Washington Mine near Courtenay on Vancouver Island (Figure 8.12), but there are many similar sites across Canada and around the world.

Figure 8.4 A rock broken by ice wedging sits in a stream in Mount Revelstoke National Park, Canada. Rocks break apart when ice expands in pre-existing cracks. *Source: Karla Panchuk (2018) CC BY 4.0.*

At many ARD sites, the pH of the runoff water is less than 4 (very acidic). Under these conditions, metals such as copper, zinc, and lead easily dissolve in water, which can be toxic to aquatic life and other organisms. For many years, the river downstream from the Mt. Washington Mine had so much dissolved copper in it that it was toxic to salmon. Remediation work has since been carried out at the mine and the situation has improved.

Exercise: Chemical Weathering

For each of the following reactions, indicate which chemical weathering process—dissolution, hydrolysis, hydration, or oxidation—is the primary mechanism.

1. Pyrite → hematite
2. Calcite → calcium and bicarbonate ions
3. Feldspar → clay
4. Olivine → serpentine
5. Pyroxene → iron oxide
6. Anhydrite → gypsum

8.3 Controls on Weathering Processes and Rates

Weathering does not happen at the same rate in all environments. The same types of weathering do not happen in all environments. There are a variety of factors that determine what kinds of weathering will occur, and how fast the processes will proceed.

Climate

Water and temperature are key factors controlling both weathering rates and the types of weathering that occur:

- Water is required for chemical weathering reactions to occur.
- Water must be present for ice wedging to happen.
- Higher temperatures speed up chemical reactions.
- Climate will determine whether water is present mostly in liquid form, solid form (ice), or both.
- Climate will determine what plant life is available to force rocks apart with their roots, and to contribute organic acids to soils to aid in chemical weathering.

This means, for example, that chemical weathering will be faster in a tropical rainforest than in the Arctic, a cold desert. It means physical weathering will be the predominant form of weathering in the Arctic.

Oxygen and Carbon Dioxide

The presence and abundance of oxygen and carbon dioxide affect chemical weathering rates. Surface environments on Earth almost all have some free oxygen available, permitting oxidation reactions to take place. Exceptions are in settings such as deep lakes or swamps where oxygen cannot easily mix into the water, and where biological processes consume the oxygen rapidly.

Carbon dioxide, which acidifies water and contributes to chemical weathering, is more concentrated in some settings than others. For example, because of the activities of organisms, soils can have very high concentrations of carbon dioxide, whereas carbon dioxide concentrations will be lower on surfaces free of soils and exposed to the atmosphere.

Minerals

The minerals making up a rock will determine what kinds of chemical weathering reactions are possible, and how rapidly chemical weathering reactions occur. Under the same conditions, dissolution of the calcite making up limestone will occur more rapidly than hydrolysis reactions happening to feldspar in granite. Quartz is very resilient to chemical weathering, and will remain long after calcite and feldspar have been

weathered away. A rock with grains cemented by calcite will weather faster than a rock with grains cemented by quartz.

In general, differences in the rates of chemical weathering among minerals can be broken down as follows:

- Minerals that weather by dissolution (e.g., halite, gypsum, calcite) are the easiest to weather.
- Silicate minerals with lower silica to oxygen ratios (e.g., silicates made of isolated silica tetrahedra or single chains) are easier to weather than silicate minerals with higher ratios (e.g., those made of silica tetrahedra arranged sheets or frameworks).
- Minerals that are by-products of chemical weathering are some of the most resistant to further chemical weathering, although they may be more prone to physical weathering (e.g., clay minerals).

Weathering Makes Weathering Go Faster

Weathering accelerates weathering. Physical weathering forms cracks and breaks rocks apart into smaller pieces. The smaller the pieces, the greater the surface area exposed to chemical weathering. When the newly exposed surfaces are exposed to chemical weathering, it weakens the rock even further, making it more susceptible to physical weathering processes.

Differential Weathering

When rocks in an outcrop weather at different rates, the result is called **differential weathering**. Differential weathering causes some beds in an outcrop to be recessed relative to the others, because beds that are slow to weather will take longer to recede than weaker beds (Figure 8.15).

Figure 8.15 Differential weathering in an outcrop along the Blaeberry River near Golden BC. The recessed beds within the outcrop are weathering faster than the surrounding beds. *Source: Karla Panchuk (2009) CC BY 4.0*

8.4 Weathering and Erosion Produce Sediments

The visible products of weathering and erosion are the unconsolidated materials that we find around us on slopes, beneath glaciers, in stream valleys, on beaches, and in deserts. The loose collection of material is referred to as **sediment**, and the individual pieces that make it up are called **clasts**. Clasts can be of any size: sand-sized and smaller (in which case they might be referred to as **particles** or **grains**), or larger than a house.

Clasts can range widely in size and shape (Figure 8.16) depending on the processes involved in making and transporting them. If and when deposits like these are turned into sedimentary rocks, the mineralogy and textures of these rocks will vary significantly. Importantly, when we describe sedimentary rocks that formed millions of years in the past, we can study the mineralogy and textures to make inferences about the conditions that existed during the deposition of the sediment, and the later burial and formation of sedimentary rock. The properties we look at are **composition**, **grain size**, **sorting**, **rounding**, and **sphericity**.

Figure 8.16 Products of weathering and erosion formed under different conditions. Source: Steven Earle (2016) CC BY 4.0

Composition

Composition refers to the mineral or minerals making up the clast. Small clasts might be single mineral grains, but larger ones can have several different mineral grains, or even several different pieces of rock within them. The composition can tell us about what rock the sediments came from, and about the geological setting from which the sediment was derived.

Not all minerals have the same hardness and resistance to weathering, so as weathering and erosion proceed, some minerals become more abundant than others within sediments. Quartz is one example of a mineral that is more abundant. It is highly resistant to weathering by weak acids or reaction with oxygen. This makes it unique among the minerals that are common in igneous rocks. Quartz is also very hard, so it is resistant to mechanical weathering.

In contrast, feldspar and iron- and magnesium-bearing minerals are not as resistant to weathering. As weathering proceeds, they are likely to be broken into small pieces and converted into clay minerals and dissolved ions. Ultimately this means that quartz, clay minerals, iron oxides, aluminum oxides, and dissolved ions are the most common products of weathering.

Grain Size

Whether a grain is large or small tells us about its journey from its source to where it was deposited. Mechanical weathering can break off large pieces from rock. Large pieces carried along by streams will bump into each other, causing smaller pieces to break off. Over time the grains get smaller

and smaller still. If we find grains that are very small, we can conclude that they travelled over a long distance.

Geologists have a specific set of definitions to describe the size of grains (Figure 8.17).

	Clast name	Diameter Range
Coarse-grained	Boulder	Larger than 256 mm
	Cobble	64 mm - 256 mm
	Pebble	2 mm - 64 mm
Medium-grained	Sand	63 µm - 2 mm
	coarse	500 µm - 2 mm
	medium	250 µm - 500 µm
	fine	63 µm - 250 µm
Fine-grained	Silt	2 µm - 63 µm
	Clay	Smaller than 2 µm

Figure 8.17 Classification of grain sizes. Silt and clay are considered fine-grained particles, sand is medium-grained, and particles larger than sand are considered coarse-grained. *Source: Karla Panchuk (2016) CC BY 4.0. Text version available online.*

The scale has some of the grain sizes listed in microns (μm). There are 1000 μm in 1 mm. The particles classified as sand are what you would intuitively think of as being sand-sized, so an easy way to remember the scale is that anything smaller than sand is fine-grained, and anything larger is coarse-grained. Fine sand grains are still easily discernible with the naked eye. Silt grains are barely discernible in rocks, and silty rocks feel gritty when rubbed. Clay grains are invisible to the naked eye, and rocks comprised of clay feel smooth when rubbed.

One other thing to notice about this scale is that the finest-grained particle is referred to as clay. While a clay-sized particle could be composed of clay minerals (and often they are), it doesn't have to be. Any particle of that size would be referred to as clay.

Grain Size and Transportation

The grain size of sediments is not just for purposes of description. It's also a valuable clue to the processes that have acted on those sediments, because the size of the clast determines how much energy is required to move it.

Whether or not a medium such as water or air has the ability to move a clast of a particular size and keep it moving depends on the velocity of the flow. For the most part, the faster the medium flows, the larger the clasts that can be moved. Figure 8.18 shows a stream bed that now contains only a trickle of water—barely enough to move particles of sand or cool puppy feet. But the velocity of water in the stream changes from season to season, as does the volume of water. All of the clasts in the stream bed were transported there by water at some point.

Figure 8.18 Ruby looks upstream in a channel near Golden BC. For much of the year the only water in the stream is the trickle in which Ruby stands, but in the spring the water can flow rapidly enough to carry boulders. *Source: Karla Panchuk (2009) CC BY 4.0*

Very fine-grained particles are the exception to rule that the larger the clasts, the faster the water that is required to transport them. Clay and silt grains stick together, requiring higher water velocities to pick them up and move them than some larger particles. Water that flows fast enough to pick up sand would not be fast enough to pick up clay.

Sorting

Weathering can break off large fragments of rock, and erosion and transport can break these fragments down to smaller and smaller sizes. The extent to which the grains in sediment differ in size is described by **sorting** (Figure 8.19, top).

Figure 8.19 Top: Sorting of grains, ranging from well sorted where the grains are similar in size, to poorly sorted, where the grains vary greatly in size. Bottom: Rounding refers to how smooth or rough the edges of a clast are. Clasts with sharp edges and corners are angular. Clasts with smooth surfaces are rounded. Clasts that fall in between are sub-angular or sub-rounded. Source: Reagan et al. (2015) CC BY 3.0

If the grains in a sample of sediment are the same size or very nearly so, the sediment is said to be well sorted. If the grains vary substantially in size, the sediment is poorly sorted. Because grains become progressively smaller as they are transported, sorting improves the further the sediments are from their source.

Rounding

Rounding refers to whether clasts have sharp edges and corners or not (Figure 8.19, bottom). If the grains are rough, with lots of edges and corners, then they are referred to as angular. Grains with smooth surfaces are rounded. Grains in between can be sub-angular or sub-rounded. The farther sediments are transported, the rounder they become.

Sphericity

Sphericity describes whether a grain is elongate or not. Grains that are longer than they are wide (like an ellipse) have low sphericity, whereas grains that have the same diameter no matter where you measure it (like a sphere) have high sphericity. In the bottom row of boxes in Figure 8.19 the grains at the top of each box exhibit high sphericity, and the grains at the bottom exhibit low sphericity. Notice that a grain can be angular but still have high sphericity. It can be rounded, but still have low sphericity. Sphericity also increases the further the sediments are from their source.

Exercise: Looking at Sand

Three samples of sand are shown below. Read the descriptions of what they contain and where they are from, then describe each sample in terms of grain size, sorting, rounding, and sphericity.

Figure 8.20 Three examples of sand grains. *Source: Steven Earle (2016) CC BY 4.0*

Sample A: Fragments of red coral, algae plates, and urchin needles from a shallow water area (~2 m depth) near a reef in Belize. The grains are between 0.1 and 1 mm across.

Sample B: Quartz and rock fragments from a glacial stream deposit near Osoyoos, BC. The grains are between 0.25 and 0.5 mm in diameter.

Sample C: Grains of olivine (green) and volcanic glass (black) from a beach on the big island of Hawai'i. The grains are approximately 1 mm in diameter.

8.5 Weathering and Soil Formation

Weathering is a key part of the process of soil formation, and soil is critical to our existence on Earth. In other words, we owe our existence to weathering, and we need to take care of the soil!

Many people refer to any loose material on Earth's surface as soil, but to scientists **soil** is the material that includes organic matter, forms within the top few tens of centimetres of the surface, and is important for sustaining plant growth.

Soil is a complex mixture of minerals (~45%), organic matter (~5%), and empty space (~50%, filled to varying degrees with air and water). The mineral content of soil varies, but is dominated by clay minerals and quartz, along with minor amounts of feldspar and small fragments of rock.

The types of weathering that take place within a region have a major influence on soil composition and texture. For example, in a warm climate where chemical weathering dominates, soils tend to be richer in clay. Soil scientists describe soil texture in terms of the relative proportions of sand, silt, and clay (Figure 8.21). Sand and silt components are dominated by quartz, with lesser amounts of feldspar and rock fragments. The clay component is dominated by clay minerals.

Figure 8.21 Soil texture classification diagram. Textures are determined by the proportions of sand-, silt-, and clay-sized grains. *Source: Mike Norton (2011) CC BY-SA 3.0*

Factors Affecting How Soil Forms

Soil forms through the mechanical and chemical weathering of rocks and sediments, and the accumulation and decay of organic matter. The factors that affect the nature of soil and the rate of its formation include:

- Climate, especially average temperature and precipitation amounts, and the consequent types of vegetation
- The parent rock or sediment that was weathered to make the soil
- The slope of the surface where soil is accumulating
- How long soil has been forming at a location

Climate

Both the mechanical breakup of rocks and the chemical weathering of minerals contribute to soil formation. The downward percolation of water brings dissolved ions and also facilitates chemical reactions. Soil forms most readily under temperate to tropical conditions, and moderate precipitation. Temperature matters because chemical weathering reactions and those facilitated by organisms proceed fastest under warm conditions, and plant growth is enhanced in warm climates. Where the climate is cooler, the rates of chemical weathering reactions decrease, and when water is frozen, may cease entirely.

Although water is needed for chemical weathering to take place, too much water can lead to soils that lack nutrients. In rain forests, for example, high rainfall contributes so much water that important nutrients are leached away, and acidic soils are left behind. In humid and poorly drained regions, swampy conditions may prevail, producing soil that is dominated by organic matter, but low in inorganic nutrients.

Too little water (e.g., in deserts and semi-deserts) limits the rate of downward chemical transport, and it also means that salts and carbonate ions dissolved in upward-moving groundwater can precipitate and build up in sediments, hindering organic activity. These soils also lack organic matter (Figure 8.22).

Parent Material

Parent material for soils can be any type of bedrock, and any type of unconsolidated sediment, such as glacial deposits and stream deposits. Soils are described as **residual soils** if they develop on bedrock, and **transported soils** if they develop on transported material such as glacial sediments. This doesn't mean that the soils themselves have been transported, but that the soil developed on unconsolidated material rather than on bedrock.

Sandy soils develop from quartz-rich parent material, such as granite, sandstone, or loose sand. Quartz-poor material, such as shale or basalt, generates soils with little sand.

Figure 8.22 Soil consisting of wind-blown silt (loess) and little organic matter in an arid part of north-eastern Washington state. *Source: Steven Earle (2016) CC BY 4.0*

Parent materials provide important nutrients to residual soils. For example, a minor constituent of granitic rocks is the calcium-phosphate mineral apatite, which is a source of the important soil nutrient phosphorus. Basaltic parent material tends to generate very fertile soils because, in addition to phosphorus, it provides significant amounts of iron, magnesium, and calcium. The iron, magnesium, and calcium come from minerals such as olivine ($(Mg,Fe)_2SiO_4$) and plagioclase feldspar ($CaAl_2Si_2O_8$) in the basalt.

Some unconsolidated materials, such as river-flood deposits, make for especially good soils because they tend to be rich in clay minerals. Clay minerals have large surface areas with negative charges that are attractive to positively charged elements like calcium, magnesium, iron, and potassium — important nutrients for plant growth.

Slope

Soil can only develop where surface materials remain in place and are not frequently washed away or lost to mass wasting (landslides). Soils cannot develop where the rate of soil formation is lower than the rate of erosion, so steep slopes tend to have little or no soil.

Time

Even under ideal conditions, soil takes thousands of years to develop. Virtually all of southern Canada was covered with glaciers up until 14,000 years ago, and most of the central and northern parts of BC, the prairies, Ontario, and Quebec were still glaciated at 12,000 years ago. Glaciers remained in the central and northern parts of Canada until around 10,000 years ago, so conditions were still not ideal for soil development even in the southern regions. This means that soils in Canada, particularly in central and northern Canada, are relatively young and not well developed.

The same applies to soils that are forming on newly created surfaces, such as recent deltas or sand bars, in areas of mass wasting, or where an area has been resurfaced by volcanic deposits.

Because soil takes so long to form, human activities that damage soils have long-term consequences for ecosystems, and for the utility of the soil for food production.

Soil Horizons

When soils form, the downward movement of clay, water, and dissolved ions can lead to the development of chemically and texturally distinct layers known as **soil horizons**. In temperate climates, soil horizons that develop are the following (Figure 8.23):

- O horizon— A layer of organic matter
- A horizon— Partially decayed organic matter mixed with mineral material
- E horizon— The eluviated (leached) layer from which some of the clay and iron have been removed to create a pale layer that may be sandier than the other layers.
- B horizon— Where clay, iron, and other elements from the overlying soil accumulate
- C horizon— Contains broken fragments of rock. Weathering of the underlying bedrock or sediments is not yet complete.

Figure 8.22 Typical horizons in a temperate soil, from Wales. *Source: Karla Panchuk (2018) CC BY 4.0. Photograph: Richard Hartnup (2005) Public Domain*

Although rare in Canada, another type of layer that develops in hot arid regions is known as **caliche** (pronounced *ca-lee-chee*). It forms from the downward (or in some cases upward) movement of calcium ions, and the precipitation of calcite within the soil. When well developed, caliche cements the surrounding material together to form a layer that has the consistency of concrete.

How Soil Is Lost

Like all geological materials, soil is subject to erosion. Under natural conditions on gentle slopes, the rate of soil formation either balances or exceeds the rate of erosion. However, human practices related to forestry and agriculture have significantly upset this balance.

Soils are held in place by vegetation. When vegetation is removed, either through cutting trees or routinely harvesting crops and tilling the soil, this protection is lost. When soil is not protected, wind and water can easily erode it away.

Water erosion is accentuated on sloped surfaces because fast-flowing water has greater eroding power than still water. Raindrops can disaggregate exposed soil particles, putting clay into suspension in the water. **Sheetwash**—unchannelled flow across a surface—carries suspended material away, and channels erode right through the soil layer, removing both fine and coarse material (Figure 8.24). Wind erosion is exacerbated by the removal of trees that act as wind breaks, and by agricultural practices that leave bare soil exposed (Figure 8.25).

Tillage is also a factor in soil erosion, especially on slopes, because each time the soil is lifted by a cultivator, it is moved a few centimetres down the slope.

Figure 8.24 Soil erosion by rain and unchannelled runoff on a field in Alberta. *Source:Alberta Agriculture and Rural Development. See Appendix C for terms of use.*

Figure 8.25 Soil erosion by wind in Alberta. *Source:Alberta Agriculture and Rural Development. See Appendix C for terms of use.*

8.6 Soils of Canada

Until the 1950s, the classification of soils in Canada was based on the system used in the United States. However, it was long recognized that the U.S system did not apply well to many parts of Canada because of climate and environmental differences. The Canadian System of Soil Classification was first outlined in 1955 and has been refined and modified numerous times since then. There are 10 orders of soil recognized in Canada (Table 8.1), and you can explore the distribution of soils using Agriculture and Agri-Food Canada's interactive map (Figure 8.26).

Table 8.1 Canadian Soil Classification System

Order	Description	Environment
Forests		
Podzolic	Well-developed A and B horizons	Coniferous forests throughout Canada
Luvisolic	Clay-rich B horizon	Northern prairies and central BC, mostly on sedimentary rocks
Brunisolic	Poorly developed or immature soil, that does not have the well-defined horizons of podsol or luvisol	Boreal-forest soils in the discontinuous permafrost areas of central and western Canada, and also in southern BC.
Grasslands		
Chernozemic	High levels of organic matter and an A horizon at least 10 cm thick	Southern prairies and parts of BC's southern interior, in areas that experience summer water deficits
Solonetzic	A clay-rich B horizon, commonly with a salt-bearing C horizon	Southern prairies, in areas that experience water deficits during the summer
Glacial and tundra		
Cryosolic	Poorly developed soil, mostly C horizon	Permafrost areas of northern Canada
Vertisolic	Clay-rich soils associated with glacial lake deposits	Southern prairies
Other		
Organic	Dominated by organic matter; mineral horizons are typically absent	Wetland areas, especially along the western edge of Hudson Bay, and in the area between the prairies and the boreal forest
Regosolic	Does not have a B horizon (i.e., no accumulation of leached minerals)	Unstable sediments including steep slopes prone to landslides, shifting sand dunes, and floodplains where sediments are frequently moved by streams
Gleysolic	Colour patterns related to the absence of oxygen	Water-saturated soils

Figure 8.26 Distribution of soil orders in Canada in Agrifood and Agriculture Canada's interactive map tool. Go to http://bit.ly/soilmap to access the interactive map. *Source: Agrifood and Agriculture Canada. Contains information licensed under the Open Government Licence – Canada. See Appendix C for terms of use.*

Note: See the Resources section at the end of this chapter for additional information on Canadian soils.

Processes of soil formation include downward transport of solid and dissolved materials, and the nature of those processes depends in large part on the climate. In Canada's predominantly cool and humid climate—characteristic of most places other than the far north— **podzolization** is the norm. This involves downward transportation of hydrogen, iron, and aluminum from the upper part of the soil profile, and accumulation of clay, iron, and aluminum in the B-horizon. Most of the **podzols, luvisols**, and **brunisols** of Canada form through various types of podzolization.

In the grasslands of the dry southern parts of the prairie provinces and in some of the drier parts of southern BC, dark brown organic-rich **chernozem** soils are dominant. In some cases, weak calcification takes place when calcium is leached from the upper layers and accumulates in the B-horizon. Development of caliche layers is rare in Canada.

Organic soils form in areas with poor drainage and a rich supply of organic matter, such as in swamps. These soils have very little mineral matter.

In the permafrost regions of the north, where glacial retreat was most recent, the time available for soil formation has been short and the rate of soil formation slow. The soils are called **cryosols** (the *cryo* prefix is used to indicate extreme cold). In permafrost areas, the freeze-thaw process churns the soil, resulting in limited soil horizon development.

Exercise: Soils of Canada

Examine Figure 8.26 showing the distribution of soils in Canada, or use the interactive map by clicking on the figure. For each of the five soils types listed below, briefly describe the distribution. Explain the distribution based on what you know about the conditions under which the soil forms and the variations in climate and vegetation related to it.

Soil Type	Distribution	Explain the Distribution
Chernozem		
Luvisol		
Podzol		
Brunisol		
Organic		

8.7 Weathering and Climate Change

Carbon cycling on Earth operates on different timescales depending on the components of the Earth system that are involved. Over the short term, biological processes are important. In particular, living organisms — mostly plants — consume carbon dioxide from the atmosphere to make their tissues. After they die, the carbon is released back into the atmosphere over years to decades as the plant matter decays.

Over the longer term, geological processes drive the carbon cycle. Geological carbon-cycle processes operate very slowly, but they affect much more of Earth's carbon than the biological component. Carbon can move from the biological cycle to the geological cycle if it is buried in sedimentary rocks. The biological carbon could be fragments of plant material or organic molecules that are preserved as coal or in organic-rich shale. It could also be calcium carbonate body parts of marine organisms that are preserved in limestone.

The geological component of the carbon cycle is shown in Figure 8.27. The various steps in the process (not necessarily in this order) are as follows:

a. Organic matter from plants is stored in peat, coal, and permafrost for thousands to millions of years.
b. Weathering of silicate minerals converts atmospheric carbon dioxide to dissolved bicarbonate, which is stored in the oceans for thousands to tens of thousands of years.
c. Dissolved carbon is converted by marine organisms to calcite, which is stored in carbonate rocks for tens of millions to hundreds of millions of years.
d. Organic carbon compounds are stored in sediments for tens of millions to hundreds of millions of years; some end up in petroleum deposits.
e. Carbon-bearing sediments are transferred to the mantle, where the carbon may be stored for tens of millions to billions of years.
f. During volcanic eruptions, carbon dioxide is released back to the atmosphere, where it is stored for years to decades.

Figure 8.27 The geological component of the carbon cycle includes: (a) organic carbon in peat, coal and permafrost, (b) weathering of silicate minerals converts atmospheric carbon dioxide to dissolved bicarbonate, (c) marine organisms convert dissolved carbon to calcium carbonate, (d) carbon compounds are stored in sediments, (e) carbon-bearing sediments are transferred to longer-term storage in the mantle, and (f) carbon dioxide is released back to atmosphere during volcanic eruptions. *Source: Steven Earle (2016) CC BY 4.0*

At some times in Earth's history, the geological carbon cycle has been balanced, with carbon being released to the atmosphere by some processes at approximately the same rate as other processes store it. Under these conditions, the climate can remain relatively stable.

At other times, the balance is upset. Prolonged periods of greater than average volcanism can cause an imbalance. The eruption of the Siberian Traps at around 250 Ma warmed the climate significantly over a few million years, leading to a mass extinction.

Mountain-building events may also cause an imbalance. The formation of the Himalaya range between about 40 Ma and 10 Ma ago exposed rocks to weathering over a large region. The over-all rate of weathering on Earth increased because the mountains were so high, and the range was so extensive. The weathering of these rocks — most importantly the hydrolysis of feldspar — consumed atmospheric carbon dioxide and transferred carbon to the oceans and to ocean-floor carbonate minerals. Decreasing carbon dioxide levels contributed to climate cooling that culminated in the Pleistocene glaciations.

Today, burning fossil fuels is causing an imbalance in the carbon cycle. Burning coal, oil, and gas releases in a geological instant carbon that was stored by the biological carbon cycle over hundreds of millions of years. Scientists who study Earth's past climate tell us that today carbon dioxide is being added to the atmosphere faster than during some of the most extreme climate change events in Earth history. Eventually, higher carbon dioxide levels will accelerate chemical weathering, and that will help to remove some of the carbon dioxide from the atmosphere. However, weathering is part of the geological carbon cycle, and operates over long timescales. If humans stopped burning all fossil fuels today, it could still take thousands of years for balance to be restored.

Chapter 8 Summary

The topics covered in this chapter can be summarized as follows:

8.1 Mechanical Weathering

Rocks weather when they are exposed to surface conditions. In most cases, conditions at Earth's surface are very different from the conditions under which the rocks formed. Mechanical weathering processes include exfoliation, freeze-thaw, salt crystallization, and the wedging effects of plant growth.

8.2 Chemical Weathering

Chemical weathering takes place when minerals within rocks are not chemically stable in their existing environment. Chemical weathering processes include hydrolysis of silicate minerals to form clay minerals, oxidation of iron in silicate and other minerals to form iron oxide minerals, and dissolution of calcite.

8.3 Controls on Weathering Processes and Rates

Chemical weathering is faster when temperatures are warmer and moisture is present. Physical weathering is more important in regions with frequent freeze-thaw cycles. Weathering rates can depend on the abundance oxygen and carbon, and will vary with the mineral composition of a rock. Weathering accelerates weathering by exposing more surface area to chemical reactions.

8.4 Weathering and Erosion Produce Sediments

Quartz grains are one of main products of weathering and erosion because quartz is resistant to chemical and mechanical weathering. Clay minerals, iron oxide and iron hydroxide minerals, aluminum hydroxide minerals, and ions in solution are common products of chemical weathering. Particles produced by weathering can be described in terms of their composition, grain size, sorting, rounding, and sphericity.

8.5 Weathering and Soil Formation

Soil is a mixture of fine mineral fragments (including quartz and clay minerals), organic matter, and empty spaces that may be partially filled with water. Soil formation is controlled by climate (especially temperature and humidity), the nature of the parent material, the slope (because soil can't accumulate on steep slopes), and the amount of time available. Typical soils have layers called horizons, which form because of differences in the conditions with depth.

8.6 Soils of Canada

Canada has a range of soil types related to our unique conditions. The main types of soil form in forested and grassland regions, but there are extensive wetlands in Canada that produce organic soils, and large areas where soil development is poor because of cold conditions.

8.7 Weathering and Climate Change

The geological component of the carbon cycle affects Earth's climate over the long term by changing atmospheric carbon dioxide levels. Carbon is added to the atmosphere during volcanic eruptions. It is extracted from the atmosphere when silicate minerals are weathered, and when it is transformed into organic matter by plants. Organic matter can be stored in soil, permafrost, and rocks. Burning of fossil fuels involves moving carbon from geological reservoirs to the atmosphere on timescales much faster than the geological carbon cycle operates.

Review Questions

1. What must happen to a body of rock before exfoliation can occur?
2. Saskatchewan's climate is consistently cold in the winter and consistently warm in the summer. What times of year would frost wedging to be an important weathering mechanism?
3. What are the products of the hydrolysis of the feldspar albite ($NaAlSi_3O_8$)?
4. Oxidation weathering of pyrite (FeS_2) can lead to acid rock drainage (ARD). What are the environmental impacts of ARD?
5. Imagine that you and a friend encounter an old graveyard on a walk through the forest. You see a granite tombstone with the date 1705 carved into it. The tombstone next to it is too badly weathered to read the date. Your friend looks at the badly weathered stone and declares, "Look how weathered this stone is! It must be from *way* before 1705." Do you agree?
6. Many sand deposits are dominated by quartz, with very little feldspar. What weathering and erosion conditions are required to get feldspar-rich sand?
7. What ultimately happens to most of the clay that forms during the hydrolysis of silicate minerals?
8. Why are the slope and the parent materials important factors in soil formation?

9. Which soil constituents move downward to produce the B-horizon of a soil?
10. What are the main processes that lead to the erosion of soils in Canada?
11. Where in Canada would you expect to find a chernozemic soil? What characteristics of this region produce this type of soil?
12. Where are brunisolic soils found in Saskatchewan?
13. Why does weathering of silicate minerals, especially feldspar, lead to consumption of atmospheric

Answers at the end of the chapter

Resources

Soils of Canada (University of Saskatchewan): http://www.soilsofcanada.ca/

Soil Classification: Soil Orders of Canada (videos about each soil order): http://soilweb.landfood.ubc.ca/classification/

References

Bienfait, B., & Ertl P. (2013). JSME: a free molecule editor in JavaScript. *Journal of Cheminformatics 5*(24). https://doi.org/10.1186/1758-2946-5-24

Reagan, M.K., Pearce, J.A., Petronotis, K., and the Expedition 352 Scientists, (2015). *Proceedings of the International Ocean Discovery Program,* Volume 352, publications.iodp.org, doi:10.14379/iodp.proc.352.102.2015

Answers to Review Questions

1. Rock must be exposed at surface. It has to be uplifted from where it formed deep in the crust, and the material on top has to be eroded.
2. Frost wedging is most effective when temperatures swing between freezing and thawing from day to day. In Saskatchewan that happens consistently in the early spring and late fall.
3. The feldspar albite ($NaAlSi_3O_8$) will be converted to a clay (such as kaolinite) and sodium ions in solution.
4. Acid rock drainage (ARD) creates acidic stream runoff. It also increases the solubility of a wide range of metals, some of which are toxic to wildlife and humans.
5. If the stones are both granite, then it would be reasonable to conclude that the badly weathered tombstone is much older than the other, because it has been exposed to weathering for much longer. On the other hand, if the badly weathered stone is a rock that is less resistant to weathering, like limestone, then the badly weathered stone could be the same age, or even younger than the granite one.
6. Feldspar-rich sand is formed where granitic rocks are being weathered and where mechanical weathering predominates over chemical weathering. For a deposit of feldspar-rich sand to be preserved, the sand must be deposited close to its source to limit the opportunities for chemical weathering.
7. Most of the clay that forms during hydrolysis of silicate minerals ends up in rivers and is washed out to the oceans. There it eventually settles to the sea floor.
8. On a steep slope, gravity will remove materials, making it unlikely for soils to accumulate. The mineral composition of the parent rock or sediment will influence the composition of the resulting soil.
9. Clay minerals and iron move downward to produce the B-horizon of a soil.

10. Wind and water are the main processes of soil erosion in Canada. Removal of vegetation makes it easier for erosion to happen.
11. Chernozemic soils are common in the southern prairies, where organic matter from grasslands is added to soils
12. Brunisolic soils are found in the northern half of Saskatchewan where forest cover is common.
13. The weathering of feldspar to clay involves the conversion of atmospheric carbon dioxide to dissolved bicarbonate, which ends up in the ocean.

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-NC-SA 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 9. Sedimentary Rocks

Adapted by Karla Panchuk from *Physical Geology* by Steven Earle

Figure 8.1 Figure 9.1 Cretaceous sedimentary rocks exposed along a road near Drumheller, Alberta, Canada. Sedimentary rocks form in layers called beds, and the planar boundaries that separate each bed are called contacts. Each bed tells a story about the conditions in which it formed. In this picture the beds are indicating that sea level repeatedly rose and fell. The black layer about halfway up the picture is a coal seam. It tells us that the environment at that time was swampy. *Source: Karla Panchuk (2008) CC BY 4.0*

Learning Objectives

After reading this chapter and answering the review questions at the end, you should be able to:

- Explain the differences between the four kinds of sedimentary rocks: clastic, chemical, biochemical, and organic.
- Describe some of the specific kinds of rocks in each of the four categories, and the depositional environments in which they form.
- Describe the various terrestrial and marine sedimentary depositional environments, and explain how the formation of sedimentary basins is related to plate tectonic processes.
- Apply your understanding of the features of sedimentary rocks, including grain characteristics, sedimentary structures, and fossils, to the interpretation of past depositional environments and climates.
- Explain what groups, formations, and members are in sedimentary rocks, and why such terminology is used.

Sedimentary Rocks Form From the Products of Weathering and Erosion

Weathering and erosion (Chapter 8) are the first two steps in the transformation of pre-existing rocks into sedimentary rocks. The remaining steps in the formation of sedimentary rocks are transportation, deposition, burial, and lithification. These steps are shown on the right-hand side of the rock cycle diagram in Figure 9.2.

Figure 9.2 The rock cycle. Processes related to sedimentary rocks are shown on the right-hand side. *Source: Steven Earle (2015) CC BY 4.0*

Transportation is the movement of sediments or dissolved ions from the site of erosion to a site of deposition. This can be by wind, flowing water, glacial ice, or mass movement down a slope. **Deposition** takes place where the conditions change enough so that the sediments can no longer be transported. This could happen if the current slows down.

Burial occurs when sediments are deposited upon existing sediments, covering and compacting them. **Lithification** is what happens when those compacted sediments become cemented together to form solid sedimentary rock. Lithification occurs at depths of hundreds to thousands of metres within Earth.

Four Types of Sedimentary Rocks

Sedimentary rocks can be divided into four main types: **clastic, chemical, biochemical,** and **organic**. **Clastic** sedimentary rocks are composed mainly of material that is transported as solid fragments (called clasts), and then cemented together by minerals that precipitated from solution. **Chemical** sedimentary rocks are composed mainly of material that is transported as ions in solution. **Biochemical** sedimentary rocks also form from ions in solution, but organisms play an important role in converting those ions into calcium carbonate or silica body parts. **Organic** sedimentary rocks contain large amounts of organic matter, such as from plant leaves and tree bark.

9.1 Clastic Sedimentary Rocks

Lithification (Figure 9.3) is the process of converting sediments into solid rock. **Compaction** is the first step. Sediments that have been deposited are buried when more and more sediments accumulate above them. The weight of the overlying sediments pushes the clasts together, closing up some of the pore spaces (the gaps between grains) and forcing them together. Pore spaces often contain water (although they can also contain air or even hydrocarbons), so the water is squeezed out.

Figure 9.3 Lithification turns sediments into solid rock. Lithification involves the compaction of sediments and then the cementation of grains by minerals that precipitate from groundwater in the spaces between these grains. *Source: Karla Panchuk (2016) CC BY 4.0*

Cementation is the next step. Groundwater flowing through the remaining pore spaces contains ions, and these ions may precipitate, leaving behind minerals in the pore spaces. These minerals bind the grains together, and are referred to collectively as **cement**. Quartz and calcite are common cement minerals, but depending on pressure, temperature, and chemical conditions, cement might also include other minerals such as hematite and clay.

Figure 9.4 shows sandstone viewed under a microscope. The grains are all quartz but they appear different shades of grey because they are being viewed through cross-polarized light. It is difficult to tell the grains from the cement in this case because both are made of quartz, but in the image on the right the more obvious grain boundaries are marked with dashed lines. Some of the cement is marked with blue shading. Using the image on the right, see if you can pick out the grain boundaries in the image on the left.

Figure 9.4 Sandstone under a microscope. Grains and cement are quartz. Left-Original image. Right-Visible grain boundaries are marked with dashed lines, and some of the cement is shaded in blue. *Source: Karla Panchuk (2018) CC BY 4.0 modified after Woudloper, Public Domain*

Types of Clastic Sedimentary Rocks

Clastic sedimentary rocks are named according to the characteristics of clasts (rock and mineral fragments) that comprise them. These characteristics include grain size, shape, and sorting. The different types of clastic sedimentary rocks are summarized in Figure 9.5.

Figure 9.5 Types of clastic sedimentary rocks. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0, Photos by James St. John and R. Weller/ Cochise College. See Appendix C for more attributions.*

Coarse-Grained Clastic Rocks

Clastic sedimentary rocks in which a significant proportion of the clasts are larger than 2 mm are known as conglomerate if the clasts are well rounded, or breccia if they are angular (Figure 9.5, top row). Conglomerates form in high-energy environments, such as fast-flowing rivers, where the particles can become rounded as they bump into each other while being carried along. Breccias typically form where the particles are not transported a significant distance, such as in alluvial fans and talus slopes.

Medium-Grained Clastic Rocks

Sandstone (Figure 9.5, middle row) is a very common sedimentary rock, and there are many different kinds of sandstone. It is worth knowing something about the different types because they are organized according to characteristics that are useful for the detective work of figuring out what conditions led to the formation of a particular sandstone. Broadly, sandstones can be divided into two groups: arenite and wacke (rhymes with tacky).

Arenite is "clean" sandstone consisting mostly of sand-sized grains and cement, with less than 15% of fine-grained silt and clay in the **matrix** (the material between the sand-sized grains). Arenites are subdivided according to what the sand-sized grains are made of (Figure 9.6). If 90% or more of the grains are quartz, then the sandstone is called a **quartz arenite** (also called a **quartz sandstone**). If more than 10% of the grains are feldspar and more of the grains are feldspar than fragments of other rocks (lithic[1] fragments) then the sandstone is called an **arkosic arenite**, or just **arkose**. If the rock has more than 10% rock fragments, and more rock fragments than feldspar, it is **lithic arenite**.

Figure 9.6 A compositional triangle for arenite sandstones, with the three most common components of sand-sized grains: quartz, feldspar, and rock fragments. Arenites have less than 15% silt or clay. *Source: Steven Earle (2015) CC BY 4.0*

Wacke is a "dirty" sandstone, containing 15-75% fine-grained particles (clay, silt) in its matrix. A wacke can have more fine-grained particles than cement in its matrix, making for a crumbly rock. Wackes are subdivided in the same way that arenites are: quartz wacke, feldspathic wacke, and lithic wacke. Another name for a lithic wacke is **greywacke**.

Figure 9.7 shows thin sections[2] (microscopic views) of quartz arenite, arkose, and lithic wacke. In the images, quartz grains are marked Q, feldspar grains are marked F, and lithic fragments are marked L. Notice the relative abundances of each component in the three types of rocks.

Figure 9.7 Photos of thin sections of three types of sandstone. Some of the minerals are labelled: Q=quartz, F=feldspar and L= lithic (rock fragments). The quartz arenite and arkose have relatively little silt/clay matrix, while the lithic wacke has abundant matrix. *Source: Steven Earle (2016) CC BY 4.0*

Fine-Grained Clastic Rocks

Rock composed of at least 75% silt- and clay-sized clasts is called **mudrock** (Figure 9.5, bottom row). If a mudrock shows evidence of fine layers (laminations) and breaks into sheets, it is called **shale**. Otherwise, it is **siltstone** (dominated by silt), **mudstone** (a mix of silt and clay), or **claystone** (dominated by clay). The fine-grained nature of mudrocks tells us that they form in very low energy environments, such as lakes, flood plains, and the deep ocean.

---

[1] "Lithic" means "rock." Lithic clasts are rock fragments (multimineralic fragments), as opposed to single-mineral fragments.

[2] Thin sections are slivers of rock sliced thinly enough so that light can pass through them, and they can be examined under a microscope.

Exercise: Classifying Sandstones

Use Figures 9.6 and 9.7 to give the appropriate name to the sandstone in each of the magnified thin sections shown below.

Figure 9.8a Sandstone 1. *Source: Steven Earle (2015) CC BY 4.0*

Sandstone 1. Rounded sand-sized grains are approximately 99% quartz and 1% feldspar. Silt and clay make up less than 2% of the rock.

Figure 9.8b Sandstone 2. *Source: Steven Earle (2015) CC BY 4.0*

Sandstone 2. Angular sand-sized grains are approximately 70% quartz, 20% lithic fragments, and 10% feldspar. Silt and clay make up ~20% of the rock.

Clastic sediments are deposited in a wide range of environments, including from melting glaciers, slope failures, rivers (both fast and slow flowing), lakes, deltas, and ocean environments (both shallow and deep). Depending on the grain size in particular, they may eventually form into rocks ranging from mudstone to breccia and conglomerate. By examining clastic sedimentary rocks it is possible to translate the classification you have just learned into an interpretation of the environment in which the rocks were deposited.

Sediment Maturity

Maturity in sediments refers to the extent to which sediment characteristics reflect prolonged weathering and transport. Prolonged weathering and transport cause clasts to become smaller, rounder, and more well-sorted. It removes minerals that are more susceptible to weathering, such as feldspar, leaving a sediment consisting predominantly of quartz or clay. On the spectrum of sediment maturity, quartz sandstone or shale would be mature sedimentary rocks, and wacke or conglomerate would be an immature rocks.

9.2 Chemical and Biochemical Sedimentary Rocks

Clastic sedimentary rocks are dominated by components that have been transported as solid clasts (clay, silt, sand, etc.). In contrast, chemical and biochemical sedimentary rocks are dominated by components that have been transported as ions in solution (e.g., Na^+, Ca^{2+}, HCO_3^-, etc.). There is some overlap between the two because almost all clastic sedimentary rocks contain cement formed from dissolved ions, and many chemical sedimentary rocks include some clasts. The difference between chemical and biochemical sedimentary rocks is that in biochemical sedimentary rocks, organisms play a role in turning the ions into sediment. This means the presence and nature of biochemical sedimentary rocks are linked to the life requirements of the organisms that comprise them. In chemical sedimentary rocks, the process is inorganic, often resulting from a body of water evaporating and concentrating the ions. It is possible for one type of sedimentary rock to form from both chemical (inorganic) and biochemical (organically mediated) processes.

Chemical and biochemical sedimentary rocks are classified based on the minerals they contain, and are frequently dominated by a single mineral. It is true that some clastic sedimentary rocks, such as quartz arenite, can also be dominated by a single mineral, but the reasons for this are different. A clastic sedimentary rock can contain whatever minerals were present in the parent rock. The minerals the clastic rock ends up containing will depend on how much "processing" the sediments undergo by physical and

chemical weathering, and transport, before the sediment was cemented. On the other hand, chemical sedimentary rocks are limited largely to those minerals that are highly soluble in water. Because mineral content is a defining characteristic of chemical and biochemical sedimentary rocks, we will use it to organize our discussion of these rocks.

Carbonate Rocks

Carbonate rocks are those in which the dominant mineral contains the carbonate anion (CO_3^{2-}). The main carbonate minerals are calcite and aragonite. Both minerals have the formula $CaCO_3$ but they have different crystal structures. A less common carbonate mineral that is still important for forming carbonate rocks is dolomite, which has the formula $CaMg(CO_3)_2$. It is similar to calcite and aragonite, except that some of the calcium is replaced with magnesium. Dolomite is more common as a replacement mineral, which has replaced calcite in carbonate rocks.

Limestone

Limestone is comprised of calcite and aragonite. It can occur as a chemical sedimentary rock, forming inorganically due to precipitation, but most limestone is biochemical in origin. In fact, limestone is by far the most common biochemical sedimentary rock.

Almost all limestone forms in marine environments, and most of that forms on the shallow continental shelves, especially in tropical regions with coral reefs. Today continental shelves are relatively narrow zones along the margins of continents, but for large parts of geologic history sea-level was much higher, and large parts of the interiors of continents were flooded.

Reefs are highly productive ecosystems populated by a wide range of organisms, many of which use calcium and bicarbonate ions from seawater to make carbonate minerals (especially calcite) for their shells and other structures. These include corals as well as green and red algae, urchins, sponges, molluscs, and crustaceans. Some of micro-organisms use $CaCO_3$ to build tiny **tests** (shells) which accumulate on the ocean floor when these organisms die. Erosion can break all of these carbonate materials apart, scattering fragments throughout surrounding region (Figure 9.9).

Figure 9.9 Various corals and green algae on a reef at Ambergris, Belize. The light-coloured sand consists of carbonate fragments eroded from the reef organisms. *Source: Steven Earle (2015) CC BY 4.0*

Figure 9.10 shows a cross-section through a typical reef environment in a tropical region (normally between 40° N and 40° S). Reefs tend to form near the edges of steep drop-offs because the reef organisms thrive on nutrient-rich upwelling currents. As the reef builds up, it is eroded by waves and currents to produce carbonate sediments that are transported into the steep offshore **fore-reef** area and the shallower inshore **back-reef** area. Reef-derived sediments are dominated by reef-type carbonate fragments of all sizes, including mud.

Figure 9.10 Cross-section through a typical tropical reef. *Source: Steven Earle (2015) CC BY 4.0*

In many such areas, carbonate-rich sediments also accumulate in quiet lagoons, where mud and mollusc-shell fragments predominate (Figure 9.11, left) or in offshore areas with strong currents, where either foraminifera tests accumulate (Figure 9.11, middle) or calcite crystallizes inorganically to form **ooids** – spheres of calcite that form in shallow tropical ocean water with strong currents (Figure 9.11, right).

Figure 9.11 Carbonate rocks and sediments. Left- Mollusc-rich limestone formed in a lagoon area at Ambergris, Belize. Middle- Foraminifera-rich sediment from a submerged carbonate sandbar near to Ambergris, Belize . Right- Ooids from a beach at Joulters Cay, Bahamas. *Sources: Left, Middle- Steven Earle (2015) CC BY 4.0; Right- Wilson44691 (2010) Public Domain*

Limestone also accumulates in deeper water, from the steady settling out of the carbonate shells of tiny organisms that lived near the ocean surface. Processes on the ocean floor cause the water in the deepest parts of the ocean to become more acidic. This puts a lower limit on how deep in the ocean calcite and aragonite can accumulate, because they dissolve under acidic conditions.

Tufa and Travertine

Calcite can form chemical sedimentary rocks on land in a number of environments. **Tufa** forms at springs. The tufa towers in Figure 9.12 formed where spring water discharged into lake water. **Travertine** (which is less porous) forms at hot springs. Similar material precipitates within limestone caves to form **speleothems** (mineral deposits in caves, Figure 9.13) such as **stalactites** and **stalagmites**.

Figure 9.12 Tufa towers (made of calcium carbonate) in Mono Lake, California. Evaporation keeps the concentration of ions in the lake very high, allowing the calcium carbonate to precipitate. *Source: Brocken Inaglory (2006) CC BY-SA 3.0*

Figure 9.13 Speleothems in Cave Nefza in Tunisia *Source: Badreddine Besbes (2015) CC BY-SA 3.0*

Dolostone

Dolostone (also referred to as dolomite) is the carbonate rock made of the mineral dolomite ($CaMg(CO_3)_2$). Dolostone is quite common (there's a whole Italian mountain range named after it), which is surprising because marine organisms do not precipitate dolomite. Dolomite forms through **dolomitization**, a process thought to involve chemical reactions between magnesium-rich water percolating through rocks, and sediments containing calcite.

Calcite and dolomite can be distinguished from one another by applying a drop of weak acid to the rock; calcite will react with weak acid, whereas dolomite will not. Also, when dolomite weathers, it tends to turn buff (tan) in colour, whereas calcite tends toward grey and white.

Chert

Chert is made of silica (SiO_2). It has the same chemical formula as quartz, but is **cryptocrystalline**, meaning that the quartz crystals comprising chert are so small that it is difficult to see them even under a microscope. Chert can be a chemical sedimentary rock, often forming as beds within limestone (Figure 9.14), or as irregular lenses or blobs (nodules). It can also be biochemical. Some tiny marine organisms (e.g., diatoms and radiolaria) make their tests from silica. When they die their tiny shells (or tests) settle slowly to the bottom of the lake or ocean, where they accumulate and are transformed into chert.

Figure 9.14 Chert (brown layers) interbedded with limestone, Triassic Quatsino Fm, Quadra Island, BC. All of the layers have been folded, and the chert, being more resistant to weathering than limestone, stands out. *Source: Steven Earle (2015) CC BY 4.0*

Banded Iron Formations (BIFs)

Some ancient chert beds — most dating to between 1800 and 2400 Ma — are also part of a rock known as a **banded iron formation (BIF)**. It is a deep sea-floor deposit of iron oxide that is a common ore of iron. These rocks consist of alternating layers of dark iron oxide minerals (magnetite and hematite) and chert stained red by hematite (Figure 9.15).

Figure 9.15 An example of a banded iron formation with dark iron oxide layers interspersed with chert stained red by hematite. This rock is 2.1 billion years old. *Source: Andre Karwath CC BY-SA (2005)*

BIFs formed before Earth's atmosphere was fully oxygenated. At that time, seawater contained abundant soluble ferrous iron (Fe^{2+}). However, once cyanobacteria began releasing oxygen into the atmosphere as a byproduct of photosynthesis, the iron in the seawater reacted with the oxygen, turning it into insoluble ferric iron (Fe^{3+}). The result was that iron oxide minerals precipitated and sank to the ocean floor. The prevalence of BIFs in rocks dating from 2400 to 1800 Ma reflects a time when free oxygen was being added to the atmosphere, but removed just as quickly by chemical reactions. After 1800 Ma, little dissolved iron was left in the oceans so no more BIFs formed.

Evaporites

In arid regions, lakes and inland seas typically have no stream outlet, and the water that flows into them is removed only by evaporation. Under these conditions, the water becomes increasingly concentrated with dissolved salts, and eventually some of these salts may reach saturation levels and start to crystallize (Figure 9.16).

Figure 9.16 Spotted Lake, near Osoyoos, BC. This photo was taken in May when the water was relatively fresh because of winter rains. By the end of the summer the surface of this lake is typically fully encrusted with salt deposits. *Source: Steven Earle (2015) CC BY 4.0*

Although all evaporite deposits are unique because of differences in the chemistry of the water, in most cases minor amounts of carbonates start to precipitate when the solution is reduced to about 50% of its original volume. Gypsum ($CaSO_4 \cdot H_2O$) precipitates at about 20% of the original volume, and halite (NaCl) precipitates at 10%. Other important evaporite minerals include sylvite (KCl) and borax ($Na_2B_4O_7 \cdot 10H_2O$). Sylvite is mined as potash at numerous locations across Saskatchewan from evaporites that formed during the Devonian (~385 Ma) when an inland sea occupied much of the region.

Exercise: Making Evaporite

This is an easy experiment that you can do at home. Pour about 50 mL (just less than 1/4 cup) of very hot water into a cup and add 2 teaspoons (10 mL) of salt. Stir until all or almost all of the salt has dissolved, then pour the salty water (leaving any undissolved salt behind) into a shallow wide dish or a small plate. Leave it to evaporate for a few days and observe the result. It may look a little like the Figure 9.17. These crystals are up to ~3 mm in diameter.

Figure 9.17 Salt crystals up to ~ 3 mm across. *Source: Steven Earle (2015) CC BY 4.0*

9.3 Organic Sedimentary Rocks

Organic sedimentary rocks are those containing large quantities of organic molecules. Organic molecules contain carbon, but in this context we are referring specifically to molecules with carbon-hydrogen bonds, such as materials from the soft tissues of plants and animals. In other words, the carbon in calcite- $CaCO_3$ wouldn't make calcite an organic mineral because it isn't bonded to hydrogen.

An important organic sedimentary rock is coal. Most coal forms in swampy land adjacent to rivers and within deltas, and where climates are humid and tropical to temperate. The vigorous growth of vegetation leads to an abundance of organic matter that accumulates within stagnant, acidic water. This limits decay and oxidation of the organic material. If this situation—where the dead organic matter is submerged in oxygen-poor water—is maintained for centuries to millennia, a thick layer of material can accumulate. Limited decay will transform this layer into **peat** (Figure 9.18a, Figure 9.19 upper left).

Figure 9.18 Formation of coal. (a) Accumulation of organic matter within a swampy area forms a layer of peat; (b) The organic matter is buried under sediment and is compressed; (c) With greater burial, lignite coal forms; (d) At even greater depths, bituminous and eventually anthracite coal form. *Source: Steven Earle (2015) CC BY 4.0*

At some point the swamp deposit is covered with more sediment — typically because a river changes its course or sea level rises (Figure 9.18b). As more sediments are added, the organic matter is compressed and heated as temperatures increase with depth. This has the effect of concentrating the carbon within the coal. The amount of heating will determine how far this process progresses.

The further the process does progress, the more the coal will go from having obvious pieces of plant material within it, to being a black, shiny mass. Low-grade **lignite** coal forms at depths between 100 m to 1,500 m and temperatures up to ~50°C (Figure 9.18c). This is still a relatively early stage in the coal formation process, so the lignite commonly displays plant fossils that have not yet been destroyed in the process of coalification (Figure 9.19 upper right).

At between 1,000 m to 5,000 m depth and temperatures up to 150°C m, **bituminous coal** forms (Figure 9.18d, 9.19 lower right). At depths beyond 5,000 m and temperatures over 150°C, **anthracite** coal forms (Figure 9.19 lower left). In fact, as temperatures rise, the lower-grade forms of coal are actually being transformed from sedimentary to metamorphic rocks.

The transition from peat to anthracite results in a progressive increase in the carbon concentration, in hardness, and in the amount of energy available to be released upon combustion.

Figure 9.19 The formation of coal begins when plant matter is prevented from decaying by accumulating in low-oxygen, acidic water. A layer of peat forms. Heating and compression of peat form lignite, bituminous coal, and finally anthracite, as pressure and temperature increases. *Source: Karla Panchuk (2017) CC BY-NC-SA 4.0. Photos by R. Weller/ Cochise College and U. S. Geological Survey. See Appendix C for more attributions and terms of use.*

9.4 Depositional Environments and Sedimentary Basins

Sediments accumulate in a wide variety of environments, both on the continents and in the oceans. Some of the more important of these environments are illustrated in Figure 9.20.

Figure 9.20 Some of the important depositional environments for sediments and sedimentary rocks. *Source: Steven Earle (2015) CC BY 4.0 view source, modified after Mike Norton (2008) CC BY-SA 2.0*

Tables 9.1 and 9.2 provide a summary of the processes and sediment types that pertain to the various depositional environments illustrated in Figure 9.19. The types of sediments that accumulate in these environments are examined in more detail in the last section of this chapter.

Table 9.1 Terrestrial Depositional Environments			
Environment	**Key Transport Processes**	**Depositional Settings**	**Typical Sediments**
Glacial	Gravity, moving ice, moving water	Valleys, plains, streams, lakes	Glacial till, gravel, sand, silt, clay
Alluvial	Gravity, moving water	Where steep-sided valleys meet plains	Coarse angular fragments
Fluvial	Moving water	Streams	Gravel, sand, silt, organic matter
Aeolian	Wind	Deserts and coastal regions	Sand, silt
Lacustrine	Moving Water	Lakes	Sand, silt, clay, organic matter
Evaporite	Still water	Lakes in arid regions	Salts, clay
Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0			

Table 9.2 Marine Depositional Environments			
Environment	**Key Transport Processes**	**Depositional Settings**	**Typical Sediments**
Deltaic	Moving water	Deltas	Sand, silt, clay, organic matter
Beach	Waves, long-shore currents	Beaches, spits, sand bars	Gravel, sand
Tidal	Tidal currents	Tidal flats	Fine-grained sand, silt, clay
Reef	Waves, tidal currents	Reefs and adjacent basins	Carbonates
Shallow marine	Waves, tidal currents	Shelves, slopes, lagoons	Carbonates in tropical climates; sand/silt/clay elsewhere.
Lagoonal	Little transportation	Lagoon bottom	Carbonates in tropical climates, silt, clay
Submarine fan	Underwater gravity flows	Continental slopes, abyssal plains	Gravel, sand, silt, clay
Deep water	Ocean currents	Deep-ocean abyssal plains	Clay, carbonate mud, silica mud
Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0			

Most of the sediments that you might see around you, including talus on steep slopes, sand bars in streams, or gravel in road cuts, will never become sedimentary rocks. This is because they have only been deposited relatively recently — perhaps a few centuries or millennia ago — and will be re-eroded before they are buried deep enough beneath other sediments to be lithified. In order for sediments to be preserved long enough to be turned into rock (a process that takes millions or tens of millions of years) they need to have been deposited in a basin in which sediments can be preserved for that long. Most such basins are formed by plate tectonic processes (Figure 9.21).

Figure 9.21 Some types of tectonically produced basins: (a) trench basin, (b) forearc basin, (c) foreland basin, and (d) rift basin. *Source: Steven Earle (2015) CC BY 4.0*

Trench basins form where a subducting oceanic plate dips beneath the overriding continental or oceanic lithosphere. They can be several kilometres deep, and in many cases, host thick sequences of sediments from nearby eroding coastal mountains. There is a well-developed trench basin off the west coast of Vancouver Island.

A **forearc basin** lies between the subduction zone and the volcanic arc, and may be formed in part by friction between the subducting plate and the overriding plate, which pulls part of the overriding plate down. The Strait of Georgia, the channel between Vancouver Island and the BC mainland, is a forearc basin.

A **foreland basin** is caused by the mass of a mountain range depressing the crust. A **rift basin** forms where continental crust is being pulled apart, and the crust on both sides the rift subsides. If rifting continues this will eventually becomes a narrow sea, and then an ocean basin. The East African rift basin represents an early stage in this process.

9.5 Sedimentary Structures and Fossils

Through careful observation over the past few centuries, geologists have discovered that the accumulation of sediments and sedimentary rocks takes place according to some important geological principles that can be summarized as follows:

- The **principle of original horizontality** states that sediments accumulate in essentially horizontal layers. The implication is that tilted sedimentary layers observed today must have been subjected to tectonic forces that tilted them.
- The **principle of superposition** states that sedimentary layers are deposited in sequence, and that unless the entire sequence has been turned over by tectonic processes, the layers at the bottom are older than those at the top.
- The **principle of inclusions** states that any rock fragments in a sedimentary layer must be older than the layer that contains them. For example, the cobbles in a conglomerate must have been formed before the conglomerate.
- The **principle of faunal succession** states that there is a well-defined order in which organisms have evolved through geological time, and therefore the identification of specific fossils in a rock can be used to determine its age.

These and other principles are discussed in more detail in Chapter 19.

In addition to these principles that apply to all sedimentary rocks, a number of other important characteristics of sedimentary processes lead to the development of distinctive sedimentary features in specific sedimentary environments. By understanding the origins of these features, we can make some very useful inferences about the processes and depositional environment that ultimately resulted in the rocks that we are studying.

Bedding refers to sedimentary layers that can be distinguished from one another on the basis of characteristics such as texture, composition, colour, or weathering characteristics (Figure 9.22). They may also be similar layers separated by **partings**, narrow regions marking weaker surfaces where erosion is enhanced. Bedding is an indication of changes in depositional processes that may be related to seasonal differences, changes in climate, changes in locations of rivers or deltas, or tectonic changes. Bedding can form in almost any depositional environment.

Figure 9.22 Beds in the Triassic Sulphur Mt. Formation near Exshaw, Alberta. Bedding is defined by differences in colour and texture, and also by partings (darker lines) between beds that may otherwise appear to be similar. *Source: Steven Earle (2015) CC BY 4.0*

Cross-bedding is bedding that contains angled layers. It forms when sediments are deposited by flowing water or wind (Figure 9.23). Cross-beds in streams tend to be on the scale of cm to tens of cm, while those in **aeolian** (wind deposited) sediments can be on the scale of metres.

Figure 9.23 Cross-bedded Jurassic Navajo Formation aeolian sandstone at Zion National Park, Utah. In most of the layers the cross-beds dip down toward the right, implying wind direction from right to left during deposition. One bed dips in the opposite direction, implying a different wind direction. *Source: Steven Earle (2015) CC BY 4.0*

Cross-beds form as sediments are deposited on the leading edge of an advancing ripple or dune. Each layer is related to a different ripple that advances in the flow direction, and is partially eroded by the following ripple (Figure 9.23). Cross-bedding is a very important sedimentary structure to recognize because it can provide information on the direction of current flows and, when analyzed in detail, on other features like the rate of flow and the amount of sediment available.

Ripples, which are associated with the formation of cross-bedding under unidirectional flow, may be preserved on the surfaces of sedimentary beds. Ripples formed in flowing water can also help to determine flow direction because they tend to have their steepest surface facing down-flow. Ripples can also form from back-and-forth flows, like at a beach, but these do not leave cross-beds, and are symmetrical, without one side steeper than the other.

Figure 9.24 Formation of cross-beds as a series of ripples or dunes that migrate with the flow. Each ripple advances forward (right to left in this view) as more sediment is deposited on its leading face. *Source: Steven Earle (2015) CC BY 4.0*

Graded bedding is characterized by a change in grain size from bottom to top within a single bed. "Normal" graded beds are coarse at the bottom and become finer toward the top (Figure 9.25), a product of deposition from a slowing current. Some graded beds are reversed (coarser at the top), and this normally results from deposition by a fast-moving debris flow. Most graded beds form in a submarine fan environment, where sediment-rich flows descend periodically from a shallow marine shelf down a slope and onto the deeper sea floor.

In a stream environment, boulders, cobbles, and pebbles can become **imbricated**, meaning that they are generally tilted in the same direction. Clasts in streams tend to tilt with their upper ends pointing downstream, because this is the most stable position with respect to the stream flow (Figure 9.26).

Mud cracks form when a shallow body of water (e.g., a tidal flat or pond), into which muddy sediments have been deposited, dries up and cracks (Figure 9.27). This happens because the clay in the upper mud layers shrinks upon drying.

The various structures described above are critical to understanding and interpreting the formation of sedimentary rocks. In addition to these structures, geologists also look very closely at sedimentary grains to determine their mineralogy or lithology (in order to make inferences about the type of source rock and the weathering processes), their degree of rounding, their sizes, and the extent to which they have been sorted by transportation and depositional processes.

Figure 9.25 Graded bedding going from pebbles at the bottom to sand at the top. *Source: Cropped from James St. John (2018) CC BY 2.0*

A Note About Fossils

Fossils are not covered in detail in this book, but they are extremely important for understanding sedimentary rocks. Fossils can be used to date sedimentary rocks, but just as importantly, they tell us a great deal about the depositional environment of the sediments and the climate at the time: they can help to differentiate marine, aquatic, and terrestrial environments; estimate the depth of the water; detect the existence of currents; and estimate average temperature and precipitation.

Figure 9.26 Imbrication of clasts in a fluvial environment. *Source: Steven Earle (2015) CC BY 4.0*

Figure 9.27 Mud cracks in a tidal flat in England. *Source: Alan Parkinson (2000) CC BY-SA 2.0*

Exercise: Interpreting Past Environments

Sedimentary rocks can tell us a great deal about the environmental conditions that existed during the time of their formation. For each of the following rocks, make some inferences about the following:

- source rock
- weathering
- sediment transportation (medium of transport, transport distance)
- depositional conditions

Quartz sandstone: no feldspar, well-sorted and well-rounded quartz grains, cross-bedded

Feldspathic sandstone and mudstone: feldspar, volcanic fragments, angular grains, repetitive graded bedding from sandstone upwards to mudstone

Conglomerate: well-rounded pebbles and cobbles of granite and basalt; imbrication

Breccia: poorly sorted, angular limestone fragments; orange-red matrix

9.6 Groups, Formations, and Members

Geologists who study sedimentary rocks need ways to divide them into manageable units, and they also need to give those units names so that they can easily be referred to and compared with other rocks deposited in other places. The International Commission on Stratigraphy (ICS) (http://www.stratigraphy.org/) has established a set of conventions for grouping, describing, and naming sedimentary rock units.

The main stratigraphic unit is a **formation**. A formation is a series of beds that is distinct from other beds above and below, and is thick enough to be shown on the geological maps that are widely used within the area in question. In most parts of the world, geological mapping is done at a relatively coarse scale, and so most formations are on the order of a few hundred metres thick. At that thickness, a typical formation would appear on a typical geological map as an area that is at least a few millimetres thick.

A series of formations can be classified together to define a **group**, which could be as much as a few thousand metres thick, and represents a series of rocks that were deposited within a single basin (or a series of related and adjacent basins) over millions to tens of millions of years.

In areas where detailed geological information is needed (for example, within a mining or petroleum district) a formation might be divided into **members**, where each member has a specific and distinctive lithology (rock type). For example, a formation that includes both shale and sandstone might be divided into members, one of which is shale, and the other sandstone. In some areas, where even more detail is required, members may be divided into beds, but this is only applicable to beds that have a special geological significance. Groups, formations, and members are typically named for the area where they are found.

The sedimentary rocks of the Nanaimo Group on Vancouver Island provide a useful example for understanding groups, formations, and members. During the latter part of the Cretaceous Period, from about 90 Ma to 65 Ma, a thick sequence of clastic rocks was deposited in a foreland basin between what is now Vancouver Island and the BC mainland (Figure 9.28). Nanaimo Group comprises a 5000 m thick sequence of conglomerate, sandstone, and mudstone layers. Coal was mined from the Nanaimo Group rocks from around 1850 to 1950 in the Nanaimo region, and is still being mined in the Campbell River area.

Figure 9.28 The distribution of the Upper Cretaceous Nanaimo Group rocks on Vancouver Island, the Gulf Islands, and in the Vancouver area. *Source: Steven Earle (2015) CC BY 4.0, modified after Mustard (1994).*

The Nanaimo Group is divided into 11 formations (Table 9.3). In general, the boundaries between formations are based on major lithological differences.

Table 9.3 Nanaimo Group Formations

Age (Ma)	Formation	Lithologies	Depositional Environmemt
~65-66	Gabriola	Sandstone with minor mudstone	Submarine fan, high energy
~66-67	Spray	Mudstone/ sandstone turbidites	Submarine fan, low energy
~67-68	Geoffrey	Sandstone and conglomerate	Submarine fan, high energy
~68-70	Northumberland	Mudstone turbidites	Submarine fan, low energy
~70	De Courcy	Sandstone	Submarine fan, high energy
~70-72	Cedar District	Mudstone turbidites	Submarine fan, low energy
~72-75	Protection	Sandstone and minor coal	Nearshore marine and onshore deltaic and fluvial
~75-80	Pender	Sandstone and minor coal	Nearshore marine and onshore deltaic and fluvial
~80	Extension	Conglomerate, with minor sandstone and some coal	Nearshore marine and onshore deltaic and fluvial
~80-85	Haslam	Mudstone and siltstone	Shallow marine
~85-90	Comox	Conglomerate, sandstone, mudstone (coal in the Campbell River area)	Nearshore fluvial and marine

Source: Steven Earle, with data from Mustard (1994)

A wide range of depositional environments existed during the accumulation of the Nanaimo Group rocks, from nearshore marine for the Comox and Haslam Formation; to fluvial and deltaic with backwater swampy environments for the coal-bearing Extension, Pender, and Protection Formations; to a deep-water submarine fan environment for the upper six formations. The differences in the depositional environments are probably a product of variations in tectonic-related uplift over time.

The five lower formations of the Nanaimo Group are all exposed in the Nanaimo area, and were well studied during the coal-mining era between 1850 and 1950. With the exception of the Haslam formation, they were divided into members, because that was useful for understanding the rocks in the areas where coal was mined.

There is much variety in the Nanaimo Group rocks, and it would take hundreds of photographs to illustrate all of the different types of rocks. Nevertheless, a few representative examples are shown in Figures 9.30-9.32.

Figure 9.30 Nanaimo Group, Spray Formation. Turbidite layers on Gabriola Island. Each turbidite set consists of a lower sandstone layer (light colour) that grades upward into siltstone, and then into mudstone. *Source: Steven Earle (2015) CC BY 4.0*

Figure 9.31 Nanaimo Group, Pender Formation. Two separate layers of fluvial sandstone with a thin (approx. 75 cm) coal seam in between. *Source: Steven Earle (2015) CC BY 4.0*

Figure 9.32 Nanaimo Group, Comox Formation. The metal object is the end of a rock hammer that is 3 cm wide. Almost all of the clasts in this view are well-rounded basalt pebbles cobbles eroded from the Triassic Karmutsen Formation which makes up a major part of Vancouver Island. *Source: Steven Earle (2015) CC BY 4.0*

Chapter 9 Summary

The topics covered in this chapter can be summarized as follows:

9.1 Clastic Sedimentary Rocks

Sedimentary clasts are classified based on their size, and variations in clast size have important implications for transportation and deposition. Clastic sedimentary rocks range from conglomerate to mudstone. Clast size, sorting, composition, and shape are important features that allow us to differentiate clastic rocks and understand the processes that took place during their deposition.

9.2 Chemical and Biochemical Sedimentary Rocks

Chemical and biochemical sedimentary rocks form from ions that were transported in solution, and then converted into minerals by chemical and/or biological processes. The most common biochemical rock, limestone, typically forms in shallow tropical marine environments, where biological activity is a very important factor. Chert and banded iron formations can be from deep-ocean environments. Evaporites form where the waters of lakes and inland seas become supersaturated due to evaporation.

9.3 Organic Sedimentary Rocks

Organic sedimentary rocks contain abundant organic carbon molecules (molecules with carbon-hydrogen bonds). An example is coal, which forms when dead plant material is preserved in stagnant swamp water, and later compressed and heated.

9.4 Depositional Environments and Sedimentary Basins

There is a wide range of depositional environments, both on land (including glaciers, lakes, and rivers) and in the ocean (including deltas, reefs, shelves, and the deep-ocean floor). In order to be preserved, sediments must accumulate in sedimentary basins, many of which form through plate-tectonic processes.

9.5 Sedimentary Structures and Fossils

The deposition of sedimentary rocks can be described in terms of a series of principles, including original horizontality, superposition, and faunal succession. Sedimentary rocks can also have distinctive structures that are important in determining their depositional environments. Fossils are useful for determining the age of a rock, the depositional environment, and the climate at the time of deposition.

9.6 Groups, Formations, and Members

Sedimentary sequences are classified into groups, formations, and members so that they can be mapped easily and without confusion.

Review Questions

1. What are the minimum and maximum sizes of sand grains?
2. The material that makes up a rock such as conglomerate cannot be deposited by a slow-flowing river. Why?
3. Describe the two main processes of lithification.
4. What is the difference between a lithic arenite and a lithic wacke?
5. How does a feldspathic arenite differ from a quartz arenite?
6. What can we say about the source area lithology, and the weathering and transportation history of a sandstone that is primarily composed of rounded quartz grains?
7. What is the original source of the carbon that is present within carbonate deposits such as limestone?
8. What long-term environmental change on Earth led to the deposition of banded iron formations?
9. Name two important terrestrial depositional environments and two important marine ones.
10. What is the origin of a foreland basin, and how does it differ from a forearc basin?
11. Explain the origins of (a) bedding, (b) cross-bedding, (c) graded bedding, and (d) mud cracks.
12. Under what conditions will reverse-graded bedding form?
13. What are the criteria for the application of a formation name to a series of sedimentary rocks?

14. Explain why some of the Nanaimo Group formations have been divided into members, while others have not.

Answers at the end of the chapter

References

Mustard, P. (1994). The Upper Cretaceous Nanaimo Group, Georgia Basin. In J. Monger (Ed.), Geology and Geological Hazards of the Vancouver Region. *Geological Survey of Canada Bulletin 481*, 27-95.

Answers to Review Questions

1. Sand grains range in size from 1/16 mm to 2 mm.
2. Conglomerate cannot be deposited by a slow-flowing river because clasts larger than 2 mm are not transported by slow-moving water.
3. Sediments are buried beneath other sediments where, because of the increased pressure, they become compacted and water is forced out from between the grains. With additional burial they are warmed to the point where cementing minerals can form between the grains (less than 200°C).
4. Lithic arenite has less than 15% silt- and clay-sized particles, while a lithic wacke has more than 15%. Both have more than 10% rock fragments and more rock fragments than feldspar.
5. Feldspathic arenite has more than 10% feldspar and more feldspar than rock fragments. Quartz arenite has less than 10% feldspar and less than 10% rock fragments. Both have less than 15% silt and clay.
6. Source area lithology: rock that contains quartz (such as granite or sandstone). Strong weathering is required to remove feldspar, and long fluvial transportation to round the grains.
7. The carbon within carbonate deposits such as limestone ultimately comes from the atmosphere.
8. Most of Earth's banded iron formations formed during the initial oxygenation of the atmosphere between 2.4 and 1.8 Ga because iron that had been soluble in the anoxic oceans became insoluble in the oxidized oceans.
9. Terrestrial depositional environments: rivers, lakes, deltas, deserts, glaciers. Marine depositional environments: continental shelves, continental slopes, deep ocean.
10. A foreland basin forms in the vicinity of a large range of mountains where the weight of the mountains depresses the crust on either side. A forearc basin lies between a subduction zone and the related volcanic arc.
11. (a) Bedding forms where there is an interruption or change in the depositional process, or a change in the composition of the material being deposited. (b) Cross-bedding forms in fluvial or aeolian environments where sand-sized sediments are being moved and ripples or dunes are present. (c) Graded bedding forms when transport energy decreases, depositing finer and finer particles. (d) Mud cracks form where fine-grained sediments (silt or clay) are allowed to dry.
12. Reverse-graded bedding forms during gravity flows, such as debris flows.
13. A formation is a series of beds that is distinct from other beds above and below it, and is thick enough to be shown on the geological maps that are widely used within the area in question.
14. The Nanaimo Group was actively mined for coal for many decades. During that time the names were given to members and individual beds that were important to the coal miners.

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-NC-SA 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 10. Metamorphism & Metamorphic Rocks

Adapted by Karla Panchuk from *Physical Geology* by Steven Earle

Figure 10.1 Grey and white striped metamorphic rocks (called gneiss) at Pemaquid Point were transformed by extreme heat and pressure during plate tectonic collisions. *Source: Karla Panchuk (2018) CC BY 4.0. Photos by Joyce McBeth (2009) CC BY 4.0. Map by Flappiefh (2013), derivative of Reisio (2005), Public Domain.*

Learning Objectives

After reading this chapter and answering the review questions at the end, you should be able to:

- Summarize the factors that influence the nature of metamorphic rocks.
- Explain how foliation forms in metamorphic rocks.
- Classify metamorphic rocks based on their texture and mineral content, and explain the origins of both.
- Describe the various settings in which metamorphic rocks are formed and explain the links between plate tectonics and metamorphism
- Describe the different types of metamorphism, including burial metamorphism, regional metamorphism, seafloor metamorphism, subduction zone metamorphism, contact metamorphism, shock metamorphism, and dynamic metamorphism.
- Explain how metamorphic facies and index minerals are used to characterize metamorphism in a region.
- Explain why fluids are important for metamorphism and describe what happens during metasomatism.

Metamorphism Occurs Between Diagenesis And Melting

Metamorphism is the change that takes place within a body of rock as a result of it being subjected to high pressure and/or high temperature. The **parent rock** or **protolith** is the rock that exists before metamorphism starts. New metamorphic rocks can form from old ones as pressure and temperature progressively increase. The term *parent rock* is typically applied to the initial unmetamorphosed rock, rather than referring to each metamorphic rock that formed as metamorphism progresses. We don't always know whether metamorphism occurred in an uninterrupted sequence or whether metamorphism stopped and started again for different reasons at different times.

Metamorphic rocks form under pressures and temperatures that are higher than those experienced by sedimentary rocks during diagenesis, but at temperatures lower than those that cause igneous rocks to melt. Given that pressure and water content affect the temperature at which rocks melt, metamorphism can occur at higher temperatures for some kinds of rocks, whereas other rocks will begin to melt under these same conditions. Metamorphic rocks can have very different mineral assemblages and textures than their parent rocks (Figure 10.2), but their over-all chemical composition usually does not change very much.

Figure 10.2 Shale is the parent rock of gneiss (pronounced "nice"). These rocks look very different, but gneiss can form when the atoms contained within the shale are re-arranged into new mineral structures. *Source: Karla Panchuk (2018) CC BY-NC-SA. Photos by R. Weller/Cochise College. See Appendix C for photo sources.*

10.2 Controls on Metamorphic Processes

The main factors that control metamorphic processes are:

- The chemical composition of the parent rock
- The temperature at which metamorphism takes place
- The pressure applied, and whether the pressure is equal in all directions or not
- The amount and type of fluid (mostly water) that is present during metamorphism
- The amount of time over which metamorphic conditions are sustained

Mineral composition

Parent rocks can be from any of the three rock types: sedimentary, igneous, or metamorphic. The critical feature of the parent rock is its mineral composition. This is because the stability of minerals—how they are influenced by changing conditions—is what determines which minerals form as metamorphism takes place. When a rock is subjected to increased temperatures and pressures, some minerals will undergo chemical reactions and turn into new minerals, while others might just change their size and shape.

Temperature

The temperature under which metamorphism occurs is a key variable in determining which metamorphic reactions happen. Mineral stability depends on temperature, pressure, and the presence of fluids. Minerals are stable over a specific range of temperatures. Quartz, for example, is stable from surface temperatures up to approximately 1800°C. If the pressure is higher, that upper limit will also be higher. If there is water present, it will be lower. Most other common minerals have upper limits between 150°C and 1000°C.

Some minerals will change their crystal structure depending on the temperature and pressure. Quartz has different polymorphs that are stable between 0°C and 1800°C. The minerals kyanite, andalusite, and sillimanite are polymorphs with the composition Al_2SiO_5. The fact that they are stable at different pressures and temperatures means that their presence can be used to determine the pressures and temperatures experienced by a metamorphic rock that contains one or more of the polymorphs (Figure 10.3).

Figure 10.3 The Al_2SiO_5 polymorphs andalusite, kyanite, and sillimanite, and their stability fields. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Photos by Rob Lavinsky/ iRocks.com (pre-2010) CC BY-SA 3.0.*

Pressure

Pressure has implications for mineral stability, and therefore the mineral content of metamorphic rocks, but it also determines the texture of metamorphic rocks. When **directed pressure** (or **directed stress**) acts on a rock, it means the stress on the rock is much greater in one direction than another. In an experiment with cylinders of modeling clay stacked in a block (Figure 10.4, top), pushing down on the clay from above resulted in higher directed pressure in the up-down direction (larger arrows; downward from pushing on the clay, and upward from the force of the table beneath the clay) than in the sideways direction, where only air pressure was acting (small arrows). The clay cylinders became elongated in the direction of least pressure.

Rocks undergo **shear stress** when forces act parallel to surfaces. In another modelling-clay experiment, applying oppositely directed forces to the top and bottom of a block of clay (Figure 10.4, bottom) caused diagonal stretching within the block. Note the change in shape of the dashed white reference circles.

In both experiments, parts of the clay became elongated in a particular direction. When mineral grains within a rock become aligned like this, it produces a fabric called **foliation**. Foliation is described in more detail later in this chapter.

Figure 10.4 Modelling clay experiments showing the effects of pressure on textures. Top- Directed pressure. Clay was set on a flat surface and pushed down on from above (large arrows). Cylinders making up the clay block became elongated in the direction of least stress. Bottom- Shear stress applied to the top and bottom of a block of clay caused the interior to stretch. Note white dashed reference circles and elongated ellipses. *Source: Karla Panchuk (2018) CC BY 4.0*

Fluids

Water is the main fluid present within rocks of the crust, and the only one considered here. The presence of water is important for two main reasons. First, water facilitates the transfer of ions between minerals and within minerals, and therefore increases the rates at which metamorphic reactions take place. This speeds the process up so metamorphism might occur more rapidly, or metamorphic processes that might not otherwise have had time to be completed are completed.

Secondly, water—especially hot water—can have elevated concentrations of dissolved substances, making it an important medium for moving ions from one place to another within the crust. Processes facilitated by hot water are called **hydrothermal** processes (*hydro* refers to water, and *thermal* refers to heat).

Time

Most metamorphic reactions occur very slowly. Estimates of the growth rates of new minerals within a rock during metamorphism suggest that new material is added to the outside of mineral crystals at a rate of approximately 1 mm per million years. Very slow reaction rates make it difficult to study metamorphic processes in a lab.

While the rate of metamorphism is slow, the tectonic processes that lead to metamorphism are also very slow, so there is a good chance that metamorphic reactions will be completed. For example, an important setting for metamorphism is many kilometres deep within the roots of mountain ranges. A mountain range takes tens of millions of years to form, and tens of millions of years more to be eroded to the extent that we can see the rocks that were metamorphosed deep beneath it.

Exercise: How Long Did It Take?

The large reddish crystals in Figure 10.5 are garnet, and the surrounding light coloured rock is dominated by muscovite mica. The Euro coin is 23 mm in diameter. Assume that the diameters of the garnets increased at a rate of 1 mm per million years. Based on the approximate average diameter of the garnets visible, estimate how long this metamorphic process might have taken.

Figure 10.5 Garnet-mica schist from the Greek island of Syros. *Source: Graeme Churchard (2005) CC BY 2.0*

10.2 Foliation and Rock Cleavage

How Foliation Develops

When a rock is acted upon by pressure that is not the same in all directions, or by shear stress (forces acting to "smear" the rock), minerals can become elongated in the direction perpendicular to the main stress. The pattern of aligned crystals that results is called foliation.

Foliation can develop in a number of ways. Minerals can deform when they are squeezed (Figure 10.6), becoming narrower in one direction and longer in another. If a rock is both heated and squeezed during metamorphism, and the temperature change is enough for new minerals to form from existing ones, the new minerals can be forced to grow longer perpendicular to the direction of squeezing (Figure 10.7). If the original rock had bedding (represented by diagonal lines in Figure 10.7, right), foliation may obscure the bedding.

Figure 10.6 Foliation that develops when minerals are squeezed and deform by lengthening in the direction perpendicular to the greatest stress (indicated by black arrows). Left- Before squeezing. Right- After squeezing. *Source: Steven Earle (2015) CC BY 4.0.*

Figure 10.7 Effects of squeezing and aligned mineral growth during metamorphism. Left- Protolith with diagonal bedding. Right- Metamorphic rock derived from the protolith. Elongated mica crystals grew perpendicular to the main stress direction. The original bedding is obscured. *Source: Steven Earle (2015) CC BY 4.0.*

This is not always the case, however. The large boulder in Figure 10.8 in has strong foliation, oriented nearly horizontally in this view, but it also has bedding still visible as dark and light bands sloping steeply down to the right.

Figure 10.8 A geologists sits on a rock that has foliation (marked by the dashed line that is nearly horizontal), and still retains evidence of the original bedding (steeply dipping dashed line). The rock has undergone a relatively low degree of metamorphism, which is why the bedding is still visible. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0.*

Foliation and Crystal Habit

Most foliation develops when new minerals are forced to grow perpendicular to the direction of greatest stress. This effect is especially strong if the new minerals grow in platy or elongated shapes. The rock in the upper left of Figure 10.9 is foliated, and the microscopic structure of the same type of foliated rock is shown in the photograph beneath it. Over all, the photomicrograph shows that the rock is dominated by elongated crystals aligned in bands running from the upper left to the lower right. The stress that produced this pattern was greatest in the direction indicated by the black arrows, at a right angle to the orientation of the minerals. The aligned minerals are mostly mica, which has a platy crystal habit, with plates stacked together like pages in a book.

Figure 10.9 A foliated metamorphic rock called phyllite (upper left). The satin sheen comes from the alignment of minerals. Lower left- a view of the same kind of rock under a microscope showing mica crystals (colourful under polarized light) aligned in bands. The region outlined in a red dashed line shows a lens of quartz crystals that do not display alignment. Upper right- stacks of platy mica crystals. Lower right- a blocky quartz crystal. *Source: Karla Panchuk (2018) CC BY-SA 4.0. See Appendix C for more attributions.*

The zone in the photomicrograph outlined with the red dashed line is different from the rest of the rock. Not only is the mineral composition different—it is quartz, not mica—but the crystals are not aligned. The quartz crystals were subjected to the same stress as the mica crystals, but because quartz grows in blocky shapes rather than elongated ones, the crystals could not be aligned in any one direction.

Even though the quartz crystals themselves are not aligned, the mass of quartz crystals forms a lens that does follow the general trend of alignment within the rock. This happens because the stress can cause some parts of the quartz crystals to dissolve, and the resulting ions flow away at right angles to the greatest stress before forming crystals again.

The effects of recrystallization in Figure 10.9 would not be visible with the unaided eye, but when larger crystals or large clasts are involved, the effects can be visible as "shadows" or "wings" around crystals and clasts. The rock in Figure 10.10 had a quartz-rich conglomerate as a parent rock. Differential stress has caused quartz pebbles within the rock to become elongated, and it has also caused wings to form around some of the pebbles (see the pebble in the dashed ellipse). The location of the wings depends on the distribution of stress on the rock (Figure 10.10, upper right).

Figure 10.10 Metaconglomerate with elongated of quartz pebbles. The pebbles have developed "wings" to varying degrees (e.g., white dashed ellipse). These are the result of quartz dissolving where stress is applied, and flowing away from the direction of maximum stress before recrystallizing (upper right sketch). *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photo by R. Weller/ Cochise College.*

Foliation Controls How Rocks Break

Foliated metamorphic rocks have elongated crystals that are oriented in a preferred direction. This forms planes of weakness, and when these rocks break, they tend to break along surfaces that parallel the orientation of the aligned minerals (Figure 10.11). Breaks along planes of weakness within a rock that are caused by foliation are referred to as **rock cleavage**, or just cleavage. This is distinct from cleavage in minerals because mineral cleavage happens between atoms within a mineral, but rock cleavage happens between minerals.

Figure 10.11 Close-up view of a metamorphic rock with aligned elongated crystals. The crystals control the shape of the break in the rock (black gap), resulting in breaks occurring along parallel surfaces. *Source: Karla Panchuk (2018) CC BY 4.0.*

The mineral alignment in the metamorphic rock called **slate** is what causes it to break into flat pieces (Figure 10.12, left), and is why slate has been used as a roofing material (Figure 10.12, right). The tendency of slate to break into flat pieces is called **slaty cleavage**. Rock cleavage is what caused the boulder in Figure 10.8 to split from bedrock in a way that left the flat upper surface upon which the geologist is sitting.

Figure 10.12 Rock cleavage in the fine-grained metamorphic rock called slate results in breaks along relatively flat surfaces (left). This is why slate has been used for roofing material (right). *Source: Left- Roger Kidd (2008) CC BY-SA 2.0. Right- Michael C. Rygel (2007) CC BY-SA 3.0.*

10.3 Classification of Metamorphic Rocks

Metamorphic rocks are broadly classified as foliated or **non-foliated**. Non-foliated metamorphic rocks do not have aligned mineral crystals. Non-foliated rocks form when pressure is uniform, or near the surface where pressure is very low. They can also form when the parent rock consists of blocky minerals such as quartz and calcite, in which individual crystals do not align because they aren't longer in any one dimension. This distinction breaks down in zones of intense deformation, where even minerals like quartz can be squeezed into long stringers, much like squeezing toothpaste out of a tube (Figure 10.13).

Figure 10.13 Rocks from the Western Carpathians mountain range without deformation (left) and after deformation (right). Scale bar: 1 mm. Left- An undeformed granitic rock containing the mica mineral biotite (Bt), plagioclase feldspar (Pl), potassium feldspar (Kfs), and quartz (Qtz). Right- A metamorphic rock (mylonite) resulting from extreme deformation of granitic rocks. Quartz crystals have been flattened and deformed. The other minerals have been crushed and deformed into a fine-grained matrix (Mtx). *Source: Farkašovský et al. (2016) CC BY-NC-ND.*

Types of Foliated Metamorphic Rocks

Four common types of foliated metamorphic rocks, listed in order of **metamorphic grade** or intensity of metamorphism are **slate**, **phyllite**, **schist** (pronounced "shist"), and **gneiss** (pronounced "nice"). Each of these has a characteristic type of foliation

Slate

Slate (Figure 10.14) forms from the low-grade metamorphism of shale. Slate has microscopic clay and mica crystals that have grown perpendicular to the maximum stress direction. Slate tends to break into flat sheets or plates, a property described as slaty cleavage.

Phyllite

Phyllite (Figure 10.15) is similar to slate, but has typically been heated to a higher temperature. As a result, the micas have grown larger. They still are not visible as individual crystals, but the larger size leads to a satiny sheen on the surface. The cleavage of phyllite is slightly wavy compared to that of slate.

Figure 10.14 Slate, a low-grade foliated metamorphic rock. Left- Slate fragments resulting from rock cleavage. Right- The same rock type in outcrop. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Photos: Left- Vincent Anciaux (2005) CC BY-SA 3.0; Right- Gretarsson (2006) CC BY-SA 3.0*

Figure 10.15 Phyllite, a fine-grained foliated metamorphic rock. Left- Hand sample showing a satin texture. Right- The same rock type in outcrop in the city of Sopron, Hungary. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Photos: Left- Chadmull (2006) Public Domain; Right- Laszlovszky András (2008) CC BY-SA 2.5*

Schist

Schist (Figure 10.16) forms at higher temperatures and pressures and exhibits mica crystals that are large enough to see without magnification. Individual crystal faces may flash when the sample is turned in the light, making the rock appear to sparkle. Other minerals such as garnet might also be visible, but it is not unusual to find that schist consists predominantly of a single mineral.

Figure 10.16 Schist, a medium- to high-grade foliated metamorphic rock. Left- Hand sample showing light reflecting off of mica crystals. Right- Close-up view of mica crystals and garnet. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photos by R. Weller/ Cochise College.*

Gneiss

Gneiss (Figure 10.17) forms at the highest pressures and temperatures, and has crystals large enough to see with the unaided eye. Gneiss features minerals that have separated into bands of different colours. The bands of colours are what define foliation within gneiss. Sometimes the bands are very obvious and continuous (Figure 10.17, upper right), but sometimes they are more like lenses (upper left). Dark bands are largely amphibole while the light-coloured bands are feldspar and quartz. Most gneiss has little or no mica because it forms at temperatures higher than those under which micas are stable.

Figure 10.17 Gneiss, a coarse-grained, high grade metamorphic rock, is characterized by colour bands. Top- Hand samples showing that colour bands can be continuous (left) or less so (right). Bottom- Gneiss in outcrop at Belteviga Bay, Norway. Notice the light and dark stripes on the rock. *Source: Karla Panchuk (2018) CC BY-SA 4.0. See Appendix C for more attributions.*

While slate and phyllite typically form only from mudrock protoliths, schist and especially gneiss can form from a variety of parent rocks, including mudrock, sandstone, conglomerate, and a range of both volcanic and intrusive igneous rocks.

Schist and gneiss can be named on the basis of important minerals that are present: a schist derived from basalt is typically rich in the mineral chlorite, so we call it chlorite schist. One derived from shale may be a muscovite-biotite schist, or just a mica schist, or if there are garnets present it might be mica-garnet schist. Similarly, gneiss that originated as basalt and is dominated by amphibole, is an amphibole gneiss or **amphibolite** (Figure 10.18).

Figure 10.18 Amphibolite in thin section (2mm field of view), derived from metamorphism of a mafic igneous rock. Green crystals are the amphibole hornblende, and colourless crystals are plagioclase feldspar. Note horizontal crystal alignment. *Source: D.J. Waters, University of Oxford*

Types of Non-foliated Metamorphic Rocks

Metamorphic rocks that form under low-pressure conditions or under the effects confining pressure, which is equal in all directions, do not become foliated. In most cases, this is because they are not buried deeply enough, and the heat for the metamorphism comes from a body of magma that has moved into the upper part of the crust. Metamorphism that happens because of proximity to magma is called **contact metamorphism**. Some examples of non-foliated metamorphic rocks are **marble**, **quartzite**, and **hornfels**.

Marble

Marble (Figure 10.19) is metamorphosed limestone. When it forms, the calcite crystals **recrystallize** (re-form into larger blocky calcite crystals), and any sedimentary textures and fossils that might have been present are destroyed. If the original limestone is pure calcite, then the marble will be white. On the other hand, if it has impurities such as clay, silica, or magnesium, the marble could be "marbled" in appearance (Figure 10.19, bottom).

Figure 10.19 Marble is a non-foliated metamorphic rock with a limestone protolith. Left- marble made of pure calcite is white. Upper right- Microscope view of calcite crystals within marble that are blocky and not aligned. Lower right- A quarry wall showing the "marbling" that results when limestone contains components other than calcite. *Source: Karla Panchuk (2018) CC BY-NC-SA. See Appendix C for more attributions.*

Quartzite

Quartzite (Figure 10.20) is metamorphosed sandstone. It is dominated by quartz, and in many cases, the original quartz grains of the sandstone are welded together with additional silica. Sandstone often contains some clay minerals, feldspar or lithic fragments, so quartzite can also contain impurities.

Figure 10.20 Quartzite is a non-foliated metamorphic rock with a sandstone protolith. Left- Quartzite from the Baraboo Range, Wisconsin. Right- Photomicrograph showing quartz grains in quartzite from the Southern Appalachians. In the upper left half of the image, blocky quartz crystals show some evidence of alignment running from the upper right to the lower left. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Photomicrograph: Geologian (2011) CC BY-SA 4.0*

Even if formed under directed pressure, quartzite is generally not foliated because quartz crystals do not normally align with the directional pressure. On the other hand, any clay present in the original sandstone is likely to be converted to mica during metamorphism, and any such mica is likely to align with the directional pressure.

Hornfels

Hornfels is another non-foliated metamorphic rock that normally forms during contact metamorphism of fine-grained rocks like mudstone or volcanic rocks. Hornfels have different elongated or platy minerals (e.g., micas, pyroxene, amphibole, and others) depending on the exact conditions and the parent rock, yet because the pressure wasn't substantially higher in any particular direction, these crystals remain randomly oriented.

The hornfels in Figure 10.21 (left) appears to have gneiss-like bands, but these actually reflect the beds of alternating sandstone and shale that were in the protolith. They are not related to alignment of crystals due to metamorphism. On the right of Figure 10.21 is a microscopic view of another sample of hornfels, also from a sedimentary protolith. The dark band at the top is from the original bedding. Here you can see that the brown mica crystals (biotite) are not aligned.

Figure 10.21 Hornfels, a non-foliated metamorphic rock formed from a fine-grained protolith. Left- Hornfels from the Novosibirsk region of Russia from a sedimentary protolith. Dark and light bands preserve the bedding of the original sedimentary rock. The rock has been recrystallized during contact metamorphism and does not display foliation. (scale in cm). Right- Hornfels in thin section from a sedimentary protolith. Note that the brown mica crystals are not aligned. The dark band at the top reflects the layering within the sedimentary parent rock, similar to the way those layers are preserved in the sample on the left. *Source: Left- Fedor (2006) Public Domain view source; Right- D.J. Waters, University of Oxford*

What Happens When Different Rocks Undergo Metamorphism?

The nature of the parent rock controls the types of metamorphic rocks that can form from it under differing metamorphic conditions (temperature, pressure, fluids). The kinds of rocks that can be expected to form at different metamorphic grades from various parent rocks are listed in Table 10.1.

Some rocks, such as granite, do not change much at the lower metamorphic grades because their minerals are still stable up to several hundred degrees. Sandstone and limestone don't change much either because their metamorphic forms (quartzite and marble, respectively) have the same mineral composition, but re-formed larger crystals.

On the other hand, some rocks can change substantially. Mudrock (e.g., shale, mudstone) can start out as slate, then progress through phyllite, schist, and gneiss, with a variety of different minerals forming along the way. Schist and gneiss can also form from sandstone, conglomerate, and a range of both volcanic and intrusive igneous rocks.

Table 10.1 A Rough Guide to the Effect of Metamorphism on Different Protoliths

Protolith	Very Low Grade 150-300°C	Low Grade 300-450°C	Medium Grade 450-550°C	High Grade Above 550°C
Mudrock	slate	phyllite	schist	gneiss
Granite	no change			granite gneiss
Basalt	chlorite schist		amphibolite (amphibole gneiss)	
Sandstone	no change	little change	quartzite	
Limestone	little change	marble		

Note: Temperature ranges are approximate

Migmatite: Both Metamorphic and Igneous

If a metamorphic rock is heated enough, it can begin to undergo partial melting in the same way that igneous rocks do. The more felsic minerals (feldspar, quartz) will melt, while the more mafic minerals (biotite, hornblende) do not. When the melt crystallizes again, the result is light-coloured igneous rock interspersed with dark-coloured metamorphic rock. This mixed rock is called **migmatite** (Figure 10.22). Note that the foliation present in the metamorphic rock is no longer present in the igneous rock. Liquids cannot support a differential stress, so when the melt crystallizes, the foliation is gone.

A fascinating characteristic of migmatites is **ptygmatic** (pronounced "tigmatic") **folding**. These are folds look like they should be impossible because they are enveloped by rock which does not display the same complex deformation (Figure 10.23). How could those wiggly folds get in there without the rest of the rock being folded in the same way?

Figure 10.22 Migmatite photographed near Geirangerfjord in Norway. *Source: Siim Sepp (2006) CC BY-SA 3.0*

Figure 10.23 Ptygmatic folding from Broken Hill, New South Wales, Australia. Ptygmatic folding happens when a stiff layer within a rock is surrounded by weaker layers. Folding causes the stiff layer to crinkle while the weaker layers deform around it. *Source: Roberto Weinberg (http://users.monash.edu.au/~weinberg)*

The answer to the ptygmatic fold mystery is that the folded layer is much stiffer than the surrounding layers. When squeezing forces act on the rock, the stiff layer buckles but the surrounding rock flows rather than buckling, because it isn't strong enough to buckle.

Exercise: Naming Metamorphic Rocks

Which metamorphic rock is described in each of the following?

1. A rock with visible minerals of mica and with small crystals of andalusite. The mica crystals are consistently parallel to one another.
2. A very hard rock with a granular appearance and a glassy lustre. There is no evidence of foliation.
3. A fine-grained rock that splits into wavy sheets. The surfaces of the sheets have a sheen to them.
4. A rock that is dominated by aligned crystals of amphibole.

10.4 Types of Metamorphism and Where They Occur

The outcome of metamorphism depends on pressure, temperature, and the abundance of fluid involved, and there are many settings with unique combinations of these factors. Some types of metamorphism are characteristic of specific plate tectonic settings, but others are not.

Burial Metamorphism

Burial metamorphism occurs when sediments are buried deeply enough that the heat and pressure cause minerals to begin to recrystallize and new minerals to grow, but does not leave the rock with a foliated appearance. As metamorphic processes go, burial metamorphism takes place at relatively low temperatures (up to ~300 °C) and pressures (100s of m depth). To the unaided eye, metamorphic changes may not be apparent at all. Contrast the rock known commercially as Black Marinace Gold Granite (Figure 10.24)—but which is in fact a metaconglomerate—with the metaconglomerate in Figure 10.10. The metaconglomerate formed through burial metamorphism does not display any of the foliation that has developed in the metaconglomerate in Figure 10.10.

Figure 10.24 Metaconglomerate formed through burial metamorphism. The pebbles in this sample are not aligned and elongated as in the metaconglomerate in Figure 10.10. *Source: James St. John (2014) CC BY 2.0*

A Note About Commercial Rock Names

Names given to rocks that are sold as building materials, especially for countertops, may not reflect the actual rock type. It is common to use the terms *granite* and *marble* to describe rocks that are neither. While these terms might not provide accurate information about the rock type, they generally do distinguish natural rock from synthetic materials. An example of a synthetic material is the one referred to as *quartz*, which includes ground-up quartz crystals as well as resin. If you happen to be in the market for stone countertops and are concerned about getting a natural product, it is best to ask lots of questions.

Regional Metamorphism

Regional metamorphism refers to large-scale metamorphism, such as what happens to continental crust along convergent tectonic margins (where plates collide). The collisions result in the formation of long mountain ranges, like those along the western coast of North America. The force of the collision causes rocks to be folded, broken, and stacked on each other, so not only is there the squeezing force from the collision, but from the weight of stacked rocks. The deeper rocks are within the stack, the higher the pressures and temperatures, and the higher the grade of metamorphism that occurs. Rocks that form from regional metamorphism are likely to be foliated because of the strong directional pressure of converging plates.

The Himalaya range is an example of where regional metamorphism is happening because two continents are colliding (Figure 10.25). Sedimentary rocks have been both thrust up to great heights—nearly 9 km above sea level—and also buried to great depths. Considering that the normal geothermal gradient (the rate of increase in temperature with depth) is around 30°C per kilometre in the crust, rock buried to 9 km below sea level in this situation could be close to 18 km below the surface of the ground, and it is reasonable to expect temperatures up to 500°C. Notice the sequence of rocks that from, beginning with slate higher up where pressures and temperatures are lower, and ending in migmatite at the bottom where temperatures are so high that some of the minerals start to melt. These rocks are all foliated because of the strong compressing force of the converging plates.

Figure 10.25 Regional metamorphism beneath a mountain range resulting from continent-continent collision. Arrows show the forces due to the collision. Dashed lines represent temperatures that would exist given a geothermal gradient of 30 °C/km. A sequence of foliated metamorphic rocks of increasing metamorphic grade forms at increasing depths within the mountains. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0*

Seafloor (Hydrothermal) Metamorphism

At an oceanic spreading ridge, recently formed oceanic crust of gabbro and basalt is slowly moving away from the plate boundary (Figure 10.26). Water within the crust is forced to rise in the area close to the source of volcanic heat, drawing in more water from further away. This eventually creates a convective system where cold seawater is drawn into the crust, heated to 200 °C to 300 °C as it passes through the crust, and then released again onto the seafloor near the ridge.

Figure 10.26 Seafloor (hydrothermal) metamorphism of ocean crustal rock on either side of a spreading ridge. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0*

The passage of this water through the oceanic crust at these temperatures promotes metamorphic reactions that change the original olivine and pyroxene minerals in the rock to chlorite ($(Mg_5Al)(AlSi_3)O_{10}(OH)_8$) and serpentine ($(Mg, Fe)_3Si_2O_5(OH)_4$). Chlorite and serpentine are both **hydrated minerals**, containing water in the form of OH in their crystal structures. When metamorphosed ocean crust is later subducted, the chlorite and serpentine are converted into new non-hydrous minerals (e.g., garnet and pyroxene) and the water that is released migrates into the overlying mantle, where it contributes to melting.

The low-grade metamorphism occurring at these relatively low pressures and temperatures can turn mafic igneous rocks in ocean crust into **greenstone** (Figure 10.27), a non-foliated metamorphic rock.

Figure 10.27 Greenstone from the metamorphism of seafloor basalt that took place 2.7 billion years ago. The sample is from the Upper Peninsula of Michigan, USA. *Source: James St. John (2012) CC BY 2.0*

Subduction Zone Metamorphism

At subduction zones, where ocean lithosphere is forced down into the hot mantle, there is a unique combination of relatively low temperatures and very high pressures. The high pressures are to be expected, given the force of collision between tectonic plates, and the increasing lithostatic pressure as the subducting slab is forced deeper and deeper into the mantle. The lower temperatures exist because even though the mantle is very hot, ocean lithosphere is relatively cool, and a poor conductor of heat. That means it will take a long time to heat up, can be several hundreds of degrees cooler than the surrounding mantle. In Figure 10.28, notice that the isotherms (lines of equal temperature, dashed lines) plunge deep into the mantle along with the subducting slab, showing that regions of relatively low temperature exist deeper in the mantle.

Figure 10.28 Regional metamorphism of oceanic crust at a subduction zone occurs at high pressure but relatively low temperatures. *Source: Steven Earle (2015) CC BY 4.0*

A special type of metamorphism takes place under these very high-pressure but relatively low-temperature conditions, producing an amphibole mineral known as **glaucophane** ($Na_2(Mg_3Al_2)Si_8O_{22}(OH)_2$). Glaucophane is blue, and the major component of a rock known as **blueschist**. If you have never seen or even heard of blueschist, that not surprising. What is surprising is that *anyone* has seen it! Most of the blueschist that forms in subduction zones continues to be subducted. It turns into **eclogite** at about 35 km depth, and then eventually sinks deep into the mantle, never to be seen again. In only a few places in the world, the subduction process was interrupted, and partially subducted blueschist returned to the surface. One such place is the area around San Francisco. The blueschist at this location is part of a set of rocks known as the Franciscan Complex (Figure 10.29).

Figure 10.29 Franciscan Complex blueschist exposed north of San Francisco. The blue colour of the rock is due to the presence of the amphibole mineral glaucophane. *Source: Steven Earle (2015) CC BY 4.0*

Contact Metamorphism

Contact metamorphism happens when a body of magma intrudes into the upper part of the crust. Heat is important in contact metamorphism, but pressure is not a key factor, so contact metamorphism produces non-foliated metamorphic rocks such as hornfels, marble, and quartzite.

Any type of magma body can lead to contact metamorphism, from a thin dyke to a large stock. The type and intensity of the metamorphism, and width of the metamorphic **aureole** that develops around the magma body, will depend on a number of factors, including the type of country rock, the temperature of the intruding body, the size of the body, and the volatile compounds within the body (Figure 10.30). A large intrusion will contain more thermal energy and will cool much more slowly than a small one, and therefore will provide a longer time and more heat for metamorphism. This will allow the heat to extend farther into the country rock, creating a larger aureole. Volatiles may exsolve from the intruding melt and travel into the country rock, facilitating heating and carrying chemical constituents from the melt into the rock. Thus, aureoles that form around "wet" intrusions tend to be larger than those forming around their dry counterparts.

Figure 10.30 Schematic cross-section of the middle and upper crust showing two magma bodies. The upper body, which has intruded into cool unmetamorphosed rock, has created a zone of contact metamorphism. The lower body is surrounded by rock that is already hot (and probably already metamorphosed), and so it does not have a significant metamorphic aureole. *Source: Steven Earle (2015) CC BY 4.0*

Contact metamorphic aureoles are typically quite small, from just a few centimetres around small dykes and sills, to as much as 100 m around a large stock. Contact metamorphism can take place over a wide range of temperatures—from around 300 °C to over 800 °C. Different minerals will form depending on the exact temperature and the nature of the country rock.

Although bodies of magma can form in a variety of settings, one place magma is produced in abundance, and where contact metamorphism can take place, is along convergent boundaries with subduction zones, where volcanic arcs form (Figure 10.31). Regional metamorphism also takes place in this setting, and because of the extra heat associated with the magmatic activity, the geothermal gradient is typically steeper in these settings (between ~40 and 50 °C/km). Under these conditions, higher grades of metamorphism can take place closer to surface than is the case in other areas.

Figure 10.31 Contact metamorphism (yellow rind) around a high-level crustal magma chamber, and regional metamorphism in a volcanic-arc related mountain range. Dashed lines show isotherms. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0*

Shock Metamorphism

When extraterrestrial objects hit Earth, the result is a shock wave. Where the object hits, pressures and temperatures become very high in a fraction of a second. A "gentle" impact can hit with 40 GPa and raise temperatures up to 500 °C. Pressures in the lower mantle start at 24 GPa (GigaPascals), and climb to 136 GPa at the core-mantle boundary, so the impact is like plunging the rock deep into the mantle and releasing it again within seconds. The sudden change associated with shock metamorphism makes it very different from other types of metamorphism that can develop over hundreds of millions of years, starting and stopping as tectonic conditions change.

Two features of shock metamorphism are **shocked quartz**, and **shatter cones**. Shocked quartz (Figure 10.32, left) refers to quartz crystals that display damage in the form of parallel lines throughout a crystal. The quartz crystal in Figure 10.32 has two sets of these lines. The lines are small amounts of glassy material within the quartz, formed from almost instantaneous melting and resolidification when the crystal was hit by a shock wave. Shatter cones are cone-shaped fractures within the rocks, also the result of a shock wave (Figure 10.32 right). The fractures are nested together like a stack of ice-cream cones.

Figure 10.32 Shock metamorphism features. Left- Shocked quartz with lines of glassy material, from the Suvasvesi South impact structure in Finland. Right- Shatter cones from the Wells Creek impact crater in the USA. *Sources: Left- Martin Schmieder CC BY 3.0 view source. Right- Zamphuor (2007) Public Domain*

Dynamic Metamorphism

Dynamic metamorphism is the result of very high shear stress, such as occurs along fault zones. Dynamic metamorphism occurs at relatively low temperatures compared to other types of metamorphism, and consists predominantly of the physical changes that happen to a rock experiencing shear stress. It affects a narrow region near the fault, and rocks nearby may appear unaffected.

Figure 10.33 Fault breccia, created when shear stress along a fault breaks up rocks. Left- close-up view of fault breccia clearly showing dark angular fragments. Right- A fault-zone containing fragments broken from the adjacent walls (dashed lines). Note that the deformation does not extend far past the margins of the fault zone. *Source: Karla Panchuk (2018) CC BY 4.0. See Appendix C for more attributions.*

At lower pressures and temperatures, dynamic metamorphism will have the effect of breaking and grinding rock, creating **cataclastic rocks** such as **fault breccia** (Figure 10.33). At higher pressures and temperatures, grains and crystals in the rock may deform without breaking into pieces (Figure 10.34, left). The outcome of prolonged dynamic metamorphism under these conditions is a rock called **mylonite**, in which crystals have been stretched into thin ribbons (Figure 10.34, right).

Figure 10.34 Mylonite, a rock formed by dynamic metamorphism. Left- An outcrop showing the early stages of mylonite development, called protomylonite. Notice that the deformation does not extend to the rock at the bottom of the photograph. Middle- Mylonite showing ribbons formed of drawn-out crystals. Right- Microscope view of mylonite with mica (colourful crystals) and quartz (grey and black crystals). This is a case where the shape of quartz crystals is controlled more by stress than by crystal habit. *Source: Karla Panchuk (2018) CC BY-SA 4.0. See Appendix C for more attributions.*

10.5 Metamorphic Facies and Index Minerals

In any given metamorphic setting there can be a variety of protolith types exposed to metamorphism. While these rocks will be exposed to the same range of pressure and temperatures conditions within that setting, the metamorphic rock that results will depend on the protolith. A convenient way to indicate the range of possible metamorphic rocks in a particular setting is to group those possibilities into **metamorphic facies**. In other words, a given metamorphic facies groups together metamorphic rocks that form under the same pressure and temperature conditions, but which have different protoliths.

Figure 10.35 shows the different metamorphic facies as patches of different colours. The axes on the diagram are temperature and depth; the depth within the Earth will determine how much pressure a rock is under, so the vertical depth axis is also a pressure axis. Each patch of colour represents a range of temperature and pressure conditions where particular types of metamorphic rocks will form. Metamorphic facies are named for rocks that form under specific conditions (e.g., eclogite facies, amphibolite facies etc.), but those names don't mean that the facies is limited to that one rock type.

Figure 10.35
Metamorphic facies and types of metamorphism shown in the context of depth and temperature. The metamorphic rocks formed from a mudrock protolith under regional metamorphism with a typical geothermal gradient are listed. Letters correspond to the types of metamorphism shown in Figure 10.36. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2016) CC BY 4.0*

Another feature to notice in the diagram are the many dashed lines. The yellow, green, and blue dashed lines represent the geothermal gradients in different environments. Recall that the geothermal gradient describes how rapidly the temperature increases with depth in Earth. In most areas (green dashed line), the rate of increase in temperature with depth is 30 °C/km. In other words, if you go 1,000 m down into a mine, the temperature will be roughly 30 °C warmer than the average temperature at the surface. In volcanic areas (yellow dashed line), the geothermal gradient is more like 40 to 50 °C/km, so the temperature rises much faster as you go down. Along subduction zones (blue dashed line), the cold ocean lithosphere keeps temperatures low, so the gradient is typically less than 10 °C/km.

The yellow, green, and blue dashed lines in Figure 10.35 tell you what metamorphic facies you will encounter for rocks from a given depth in that particular environment. A depth of 15 km in a volcanic region falls in the amphibolite facies. Under more typical conditions, this is the greenschist facies, and in a subduction zone it is the blueschist facies. You can make the connection more directly between the metamorphic facies and the types of metamorphism discussed in the previous section by matching up the letters *a* through *e* in Figure 10.35 with the labels in Figure 10.36.

Figure 10.36 Environments of metamorphism in the context of plate tectonics: (a) regional metamorphism related to mountain building at a continent-continent convergent boundary, (b) seafloor (hydrothermal) metamorphism of oceanic crust in the area on either side of a spreading ridge, (c) metamorphism of oceanic crustal rocks within a subduction zone, (d) contact metamorphism adjacent to a magma body at a high level in the crust, and (e) regional metamorphism related to mountain building at a convergent boundary. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0.*

One other line to notice in Figure 10.35 is the red dashed line on the right-hand side of the figure. This line represents temperatures and pressures where granite will begin to melt if there is water present. Migmatite is to the right of the line because it forms when some of the minerals in a metamorphic rock begin to melt, and then cool and crystallize again.

Exercise: Metamorphic Rocks in Areas with Higher Geothermal Gradients

Figure 10.35 shows the types of rock that might form from mudrock at various points along the curve of the "typical" geothermal gradient (dotted green line). Looking at the geothermal gradient for volcanic regions (dotted yellow line), estimate the depths at which you would expect to find each of those rocks forming from a mudrock parent.

Index Minerals

Some common minerals in metamorphic rocks are shown in Figure 10.37, arranged in order of the temperature ranges within which they tend to be stable. The upper and lower limits of the ranges are intentionally vague because these limits depend on a number of different factors, such as the pressure, the amount of water present, and the overall composition of the rock.

Figure 10.37 Metamorphic index minerals and approximate temperature ranges. *Source: Steven Earle (2015) CC BY 4.0*

Even though the limits of the stability ranges are vague, the stability range of each mineral is still small enough that the minerals can be used as markers for those metamorphic conditions. Minerals that make good markers of specific ranges of metamorphic conditions are called **index minerals**.

The Meguma Terrane of Nova Scotia: An Example of How to Use Index Minerals

The southern and southwestern parts of Nova Scotia were regionally metamorphosed during the Devonian Acadian Orogeny (around 400 Ma), when a relatively small continental block—the Meguma Terrane

(Figure 10.38, top)—collided with the existing eastern margin of North America. The clastic sedimentary rocks within this terrane were variably metamorphosed.

Figure 10.38 Regional metamorphic zones in the Meguma Terrane of southwestern Nova Scotia. Top- Map of metamorphic zones. Bottom- Stability ranges for minerals within the Meguma Terrane. *Source: Karla Panchuk (2017) CC BY 4.0. Modified after Steven Earle (2015) CC BY 4.0, Keppie & Muecke (1979) and White & Barr (2012).*

Index minerals have been used to map areas of higher or lower metamorphic intensity, called **metamorphic zones**. A metamorphic zone is a region bounded by the first appearance of an index mineral. In the Meguma Terrane, the biotite zone (Figure 10.38, darker green) begins in the east and north with the first appearance of biotite. The biotite zone ends toward the south and west where garnet first appears. Because index minerals can have overlapping stability conditions, a lower-intensity index mineral can still be present in a higher-intensity metamorphic zone.

Knowledge of metamorphic zones makes it possible to draw conclusions about the geological conditions in which metamorphic rocks formed. The highest-intensity metamorphism—the sillimanite zone—is in the southwest. Progressively lower grades of metamorphism exist toward the east and north. The rocks of the sillimanite zone were likely heated to over 700 °C, and therefore must have been buried to depths between 20 km and 25 km. The surrounding lower-grade rocks were not buried as deeply, and the rocks within the peripheral chlorite zone were likely not buried to more than about 5 km depth.

A probable explanation for this pattern is that the area with the highest-grade rocks was buried beneath the central part of a mountain range formed by the collision of the Meguma Terrane with North America. The collision caused rocks to be folded, and to be faulted and stacked on top of each other. These mountain-building processes thickened Earth's crust, and increased its mass locally as the mountains grew. The increased mass of the growing mountains caused the lithosphere to float lower in the mantle (Figure 10.39, left). As the mountains were eventually eroded over tens of millions of years, the crust floated higher and higher in the mantle, and erosion exposed metamorphic rocks that were deep within the mountains (Figure 10.39, right).

Figure 10.39 Schematic cross-section through the Meguma Terrane. Left- Metamorphic zones and temperatures when mountain-building processes thickened the crust. Right- The mountains have been eroded, exposing metamorphic rocks that formed deep within the mountains. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0*

Building a narrative for the metamorphism in Nova Scotia's Meguma Terrane is just one example of how index minerals can be used.

Exercise: Scottish Metamorphic Zones

The map in Figure 10.40 shows part of western Scotland between the Great Glen Fault and the Highland Boundary Fault. The shaded areas are metamorphic rock, and the three metamorphic zones represented are garnet, chlorite, and biotite.

1. Label the three coloured areas of the map with the appropriate zone names (garnet, chlorite, and biotite).
2. Indicate which part of the region was likely to have been buried the deepest during metamorphism.

British Geologist George Barrow studied this area in the 1890s and was the first person anywhere to map metamorphic zones based on their mineral assemblages. This pattern of metamorphism is sometimes referred to as Barrovian metamorphism.

Figure 10.40 Metamorphic zones in Barrovian metamorphism. *Source: Steven Earle (2015) CC BY 4.0*

10.6 Metamorphic Hydrothermal Processes and Metasomatism

The heat from a body of magma in the upper crust can create a very dynamic situation with geologically interesting and economically important implications. In the simplest cases, water does not play a big role, and the main process is heat transfer from the pluton to the surrounding rock, creating a zone of contact metamorphism (Figure 10.41a). In many cases, however, water is released from the magma body as crystallization takes place, and this water is dispersed along fractures in the surrounding rock (Figure 10.41b). The water released from a magma chamber is typically rich in dissolved minerals. As this water cools, it interacts with the surrounding rocks, changing both the chemistry of the water and the chemistry of the rocks. This can cause minerals to precipitate from the water. Minerals can also precipitate if the water boils because of a drop in pressure. The precipitated minerals form veins within fractures in the surrounding rock. Quartz veins are commonly formed in this situation, and can include other minerals such as pyrite, hematite, calcite, and even silver and gold.

Figure 10.41 Metamorphism and alteration around a pluton in the upper crust. (a) Thermal metamorphism only (within the purple zone); (b) Thermal metamorphism plus veining (white) related to dispersal of magmatic fluids into the overlying rock; (c) Thermal metamorphism plus veining from magmatic fluids plus alteration and possible formation of metallic minerals (hatched yellow areas) from convection of groundwater. *Source: Steven Earle (2015) CC BY 4.0*

Heat from the magma body will cause surrounding groundwater to expand and then rise toward the surface. In some cases, this may initiate a convection system where groundwater circulates past the pluton. Such a system could operate for thousands of years, resulting in the circulation of millions of tonnes of groundwater from the surrounding region past the pluton.

Hot water circulating through the rocks and interacting chemically with them can lead to significant changes in the mineralogy of the rock, including alteration of feldspars to clays, and deposition of quartz, calcite, and other minerals in fractures and other open spaces (Figure 10.42). Chemical change in rocks due to interaction with hot water is called **hydrothermal alteration**.

Figure 10.42 White veins of calcite in limestone of the Comox Formation, Nanaimo BC. Quarter for scale. *Source: Steven Earle (2016) CC BY 4.0*

Metamorphic reactions involve the release of fluids as minerals change, and chemical reactions with locally-derived fluids. However, if a large amount of externally-derived fluid—such that supplied by magma—is flushed through the system at the high pressures and temperatures characteristic of metamorphism, it can substantially alter the chemical composition of the rock. This type of hydrothermal alteration is called **metasomatism**.

A special type of metasomatism takes place where a hot pluton intrudes into carbonate rock such as limestone. Magmatic fluids rich in silica, calcium, magnesium, iron, and other elements can dramatically change the chemistry of the limestone, forming minerals that would not normally exist in either the igneous rock or limestone. A rock called **skarn** results, containing minerals such as garnet, epidote, magnetite, and pyroxene, among others (Figure 10.43).

Figure 10.43 Skarn from Mount Monzoni, Northern Italy, with recrystallized calcite (blue), garnet (brown), and pyroxene (green). The rock is 6 cm across. *Source: Siim Sepp (2012) CC BY-SA 3.0*

Exercise: Contact Metamorphism and Metasomatism

A pluton that has intruded into a series of sedimentary rocks, including sandstone, mudstone, and limestone (Figure 10.44). What types of metamorphic rocks would you expect to see at locations a, b, and c?

Figure 10.44 Contact metamorphism and metasomatism of sedimentary rocks. *Source: Steven Earle (2015) CC BY 4.0*

Chapter 10 Summary

The topics covered in this chapter can be summarized as follows:

10.1 Controls on Metamorphic Processes

Metamorphism is controlled by five main factors: the composition of the parent rock, the temperature to which the rock is heated, the amount and direction of pressure, the volumes and compositions of fluids that are present, and the amount of time available for metamorphic reactions to take place.

10.2 Foliation and Rock Cleavage

When the pressure acting on a rock is not uniform in all directions, foliation can develop. Foliation may occur in the form of platy or elongated mineral crystals that have grown at right angles to the maximum pressure, or it may develop when crystals or clasts within a rock are deformed. Foliation causes crystals or clasts within a rock to become aligned. When metamorphic rocks break parallel to the direction of foliation, rock cleavage results.

10.3 Classification of Metamorphic Rocks

Metamorphic rocks are classified on the basis of texture and mineral composition. Foliation is a key feature of metamorphic rocks formed under directed pressure; foliated metamorphic rocks include slate, phyllite, schist, and gneiss. Metamorphic rocks formed in environments without strong directed pressure include hornfels, marble, and quartzite.

10.4 Types of Metamorphism and Where They Occur

Almost all regions that experience metamorphism are being acted upon by plate-tectonic processes. Oceanic crustal rock can be metamorphosed near the spreading ridge where it was formed. Regional metamorphism takes place in areas where mountain ranges are forming, which are most common at convergent boundaries. Contact metamorphism takes place around magma bodies in the crust, which are also most common above convergent boundaries. Shock metamorphism happens when extraterrestrial bodies impact Earth, and is unusual among metamorphic processes because it occurs in seconds or minutes, rather than taking millions of years. Dynamic metamorphism occurs when shear stress is applied to rocks, such as along faults.

10.5 Metamorphic Facies and Index Minerals

Metamorphic facies are groups of metamorphic rocks that form under the same range of pressure and temperature conditions, but from different parent rocks. Geologists use index minerals such as chlorite,

garnet, andalusite, and sillimanite to identify metamorphic zones. Index minerals tell us about the pressure and temperature conditions under which metamorphic rocks formed.

10.6 Metamorphic Hydrothermal Processes and Metasomatism

Contact metamorphism takes place around magma bodies that have intruded into cool rocks in the crust. Heat from magma is transferred to the surrounding country rock, resulting in mineralogical and textural changes. Hot water from a cooling body of magma, or from convection of groundwater driven by the heat of the pluton, can lead to hydrothermal alteration. When large volumes of fluid are flushed through rocks experiencing metamorphic pressures and temperatures, metasomatism results. Metasomatism can cause valuable metals to accumulate in the surrounding rocks.

Review Questions

1. What are the two main agents of metamorphism, and what are their respective roles in producing metamorphic rocks?
2. What types of metamorphic rocks will form if a mudrock experiences very low, low, medium, and high-grade metamorphism?
3. Why doesn't granite change very much at lower metamorphic grades?
4. Describe the main process of foliation development in a metamorphic rock such as schist.
5. What process contributes to metamorphism of oceanic crust at a spreading ridge?
6. How do variations in the geothermal gradient affect the depth at which different metamorphic rocks form?
7. Blueschist metamorphism takes place within subduction zones. What are the particular temperature and pressure characteristics of this geological setting?
8. Rearrange the following minerals in order of increasing metamorphic grade: biotite, garnet, sillimanite, chlorite.
9. What is the role of magmatic fluids in the metamorphism that takes place adjacent to a pluton?
10. How does metasomatism differ from regional metamorphism?
11. How does the presence of a hot pluton contribute to metasomatism?
12. What determines whether metasomatism will produce skarn?
13. For each of the following metamorphic rocks, indicate the likely parent rock and the grade and/or type of metamorphism: chlorite schist, slate, mica-garnet schist, amphibolite, marble.

Answers at the end of the chapter

References

Bucher, K., & Grapes, R. (2011) *Petrogenesis of Metamorphic Rocks, 8th Edition*. Springer.

French, B.M. (1998). *Traces of Catastrophe: A Handbook of Shock-Metamorphic Effects in Terrestrial Meteorite Impact Structures*. Houston, TX: Lunar and Planetary Institute

Farkašovský, R., Bónová, K., & Košuth, M. (2016). Microstructural, modal and geochemical changes as a result of granodiorite mylonitisation – a case study from the Rolovská shear zone (Čierna hora Mts, Western Carpathians, Slovakia). *Geologos 22*(3), 171-190. doi: 10.1515/logos-2016-0019

Keppie, D., & Muecke, G. (1979). *Metamorphic map of Nova Scotia*. (Nova Scotia Department of Mines and Energy, Map 1979-006).

White, C. E., & Barr, S. M. (2012) Meguma Terrane Revisted: Stratigraphy, Metamorphism, Paleontology and Provenance. *Geoscience Canada 39*(1).
https://journals.lib.unb.ca/index.php/GC/article/view/19450/21005

Answers to Review Questions

1. Heat and pressure are the main agents of metamorphism. Heat leads to mineralogical changes in the rock. Pressure also influences those mineralogical changes, while directed pressure (greater pressure in one direction) leads to foliation.
2. Very low grade: slate; low grade: phyllite; medium grade: schist; high grade: gneiss.
3. Granite remains largely unchanged at lower metamorphic grades because its minerals are still stable at those lower temperatures.
4. Foliation develops in schist when new platy minerals grow with their longest dimension at a right angle to the direction of greatest pressure.
5. At a spreading ridge the heat from volcanism leads to the development of a groundwater convection system in the rock of the oceanic crust. Heated water rises in the hot regions and is expelled into the ocean, while cold ocean water is drawn into the crust to replace it. The heated water leads to the conversion of olivine and pyroxene into chlorite and serpentine.
6. The geothermal gradient varies as a function of tectonic setting, being greatest in volcanic regions and lowest along subduction zones. As a result the depth at which specific metamorphic grades is achieved will vary: the depth will be greater when the gradient is lower.
7. The geothermal gradient is low within subduction zones because the cold subducting oceanic crust takes a long time to heat up. Pressure increases with depth at the normal rate, but temperature does not.
8. Order of increasing metamorphic grade: chlorite, biotite, garnet, sillimanite.
9. Water from any source facilitates metamorphism. Magmatic fluids typically contain dissolved ions at higher concentrations than in regular groundwater (especially copper, zinc, silver, gold, lithium, beryllium, boron and fluorine), leading to the formation of a unique set of minerals.
10. Metasomatism involves fluids from magmatic or groundwater sources that play an important role in transporting ions into the system, and leading to the formation of new minerals. Regional metamorphism takes place over a larger area, depends more on plate tectonic conditions, and does not involve flushing the system with large amounts of fluid.
11. A hot pluton heats the surrounding water, causing groundwater to convect. This can result a great deal of water, in some cases with elevated levels of specific ions, passing through the rock. Water from magma within the pluton also contributes to metasomatism.
12. Limestone must be present to produce skarn.

13. Parent rocks and metamorphic grades and types:

Metamorphic Rock	Likely Parent Rock	Grade and/or Type of Metamorphism
Chlorite schist	A rock enriched in ferromagnesian minerals, such as basalt	Low-grade regional metamorphism
Slate	Mudrock (shale, mudstone)	Very low grade regional metamorphism
Mica-garnet schist	A rock that is rich in aluminum, which includes most clay-bearing rocks	Medium-grade regional metamorphism
Amphibolite	A rock enriched in ferromagnesian minerals, such as basalt	Medium- to high-grade regional metamorphism
Marble	Limestone or dolomite	Regional or contact metamorphism

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-NC-SA 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 11. Volcanism

Adapted by Karla Panchuk from *Physical Geology* by Steven Earle

Figure 11.1 Mt. Garibaldi (in the background), near Squamish BC, is one of Canada's most recently active volcanoes, last erupted approximately 10,000 years ago. It is also one of the tallest, at 2,678 m in height. *Source: Karla Panchuk (2017) CC BY-SA 4.0. Photograph: Michael Scheltgen (2006) CC BY 2.0. See Appendix C for more attributions.*

Learning Objectives

After reading this chapter and answering the *Review Questions* at the end, you should be able to:

- Explain what a volcano is.
- Describe the different kinds of materials produced by volcanoes.
- Describe the structures of shield volcanoes, composite volcanoes, and cinder cones.
- Explain how the style of a volcanic eruption is related to magma composition.
- Describe the role of plate tectonics in volcanism and magma formation.
- Summarize the hazards that volcanic eruptions pose to people and infrastructure.
- Describe how volcanoes are monitored, and the signals that indicate a volcano could be ready to erupt.
- Provide an overview of Canadian volcanic activity.

Why Study Volcanoes?

Volcanoes are awe-inspiring natural events. They have instilled fear and fascination with their red-hot lava flows, and cataclysmic explosions. In his painting *The Eruption of Vesuvius* (Figure 11.2), Pierre-Jacques Volaire captured the stunning spectacle of the eruption on Mt. Vesuvius on 14 May 1771. He also captured some stunningly casual spectating being done by tourists and their dog (lower left).

Figure 11.2 *The Eruption of Vesuvius*, by Pierre-Jacques Volaire (1771). Public Domain.

As Volaire's painting suggests, curiosity alone would be enough to make people want to learn why volcanoes happen and how they work. However, there are other reasons why we should know more about volcanoes. One reason is that studying volcanoes helps us understand the evolution of the Earth system- not just Earth's geological features, but past changes in climate, and even the causes of mass extinctions. Another reason is the critical need to study the hazards posed by volcanoes to people and infrastructure. Over the past few decades, volcanologists have made great strides in their ability to forecast volcanic eruptions and predict the consequences, saving thousands of lives.

11.1 What Is A Volcano?

A **volcano** is a location where molten rock flows out, or **erupts**, onto Earth's surface as **lava**. Volcanic eruptions can happen on land or underwater. Some volcanic eruptions flow from mountains (such as Mount Garibaldi in Figure 11.1), but others do not. **Fissure eruptions** (Figure 11.3) are volcanic eruptions flowing from long cracks in the Earth.

Figure 11.3 Kamoamoa fissure eruption on the flanks of the Hawai'ian Kīlauea Volcano in March of 2011. *Source: Karla Panchuk (2017) CC BY-SA 4.0. Photograph: U. S. Geological Survey (2011) Public Domain. See Appendix C for more attributions.*

Volcano Anatomy

The main parts of a volcano are shown in Figure 11.4. When volcanoes erupt, magma moves upward from a **magma chamber** and into a **vent** or conduit. It flows out from a crater at the top, or sometimes emerges at a secondary site on the side of the volcano resulting in a **flank eruption**. Erupted materials accumulate around the vent forming a volcanic mountain. The accumulated material might consist of layers of solidified lava, called **lava flows**, but it might also include fragments of various sizes that have been thrown from the volcano.

Figure 11.4 The parts of a volcano (not to scale). *Source: Karla Panchuk (2017) CC BY 4.0*

Crater or Caldera?

A **crater** is the basin above a volcano's vent. Craters have diameters on the scale of 10s to 100s of metres. A **caldera** is a bowl-shaped structure that resembles a crater, but it is much larger (km in scale) and forms when a volcano collapses in on itself. The process is illustrated in Figure 11.5, going from left to right.

It begins when an eruption occurs, and the magma chamber beneath the volcano is drained. If a significant part of a volcano's mass is supported by magma within the chamber, then depleting the magma also reduces the support for the volcano. The loss of support causes part of the volcano to collapse into the void in the magma chamber, leaving behind a broad basin rimmed by the remnants of the volcano. Over time, the basin can fill with water. If there is still activity within the magma chamber, magma may force its way upward again, causing the floor of the caldera to be lifted, or erupting to form a new volcano within the caldera.

Figure 11.5 Formation of a caldera. Calderas are the result of a volcano collapsing into a drained magma chamber.
Source: Karla Panchuk (2017) CC BY 4.0. Modified after U. S. Geological Survey (2002) Public Domain

The island of Santorini (Figure 11.6) is an example of a caldera. The island itself is the rim of the caldera, and the bay is the flooded basin. The two small islands in the middle of the bay formed from magma refilling the chamber that feeds the volcano, as in the far right of Figure 11.5. The caldera formed after an enormous eruption between 1627 and 1600 BCE[1]. The eruption is thought to have contributed to the downfall of the Minoan civilization, and some speculate that it might also be the source of the myth of Atlantis, a story about a lost continent that sank beneath the sea after a natural disaster.

11.2 Materials Produced by Volcanic Eruptions

Volcanic eruptions produce three types of materials: gas, lava, and fragmented debris called **tephra**.

Volcanic Gas

Magma contains gas. At high pressures, the gases are dissolved within magma. However, if the pressure decreases, the gas comes out of solution, forming bubbles. This process is analogous to what happens when a pop bottle is opened. Pop is bottled under pressure, forcing carbon dioxide gas to dissolve into the fluid. As a result, a bottle of pop that you find on the supermarket shelf will have few to no bubbles. If you open the bottle, you decrease the pressure within it. The pop will begin to fizz as carbon dioxide gas comes out of solution and forms bubbles.

[1] Friedrich, W. L., Kromer, B., Friedrich, M., Heinemeier, J., Pfeiffer, T., & Talamo, S. (2006). Santorini Eruption Radiocarbon Dated 1627-1600 B.C. Science (312)5773, 548. doi: 10.1126/science.1125087

Figure 11.6 The Greek Island of Santorini. Left: Aerial view of the island forming a ring around a flooded caldera. Right: A view from the rim of the caldera. The other side of the rim is visible in the distance. *Source: Karla Panchuk (2017) CC BY-SA 4.0. Satellite image: NASA/GSFC/MITI/ERSDAC/JAROS, and U.S./Japan ASTER Science Team (2000) Public Domain; Caldera photograph: Klearchos Kapoutsis (2010) CC BY 2.0; See Appendix C for more attributions.*

The main component of volcanic gas emissions is water vapour, followed by carbon dioxide (CO_2), sulphur dioxide (SO_2), and hydrogen sulphide (H_2S).

Volcanoes release gases when erupt, and through openings called **fumaroles** (Figure 11.7). They can also release gas into soil and groundwater.

Figure 11.7 A fumarole at Pu‘u ‘Ō‘ō Crater, Hawaii. The yellow crust along the margin of the fumarole is made of sulphur crystals. The crystals form when sulphur vapour cools as it is released from the fumarole. *Source: U. S. Geological Survey (2016) Public Domain*

Lava

The ease with which lava flows and the structures it forms depend on how much silica and gas the lava contains. The more silica, the more **polymerization** (formation of long molecules) occurs, stiffening the lava. The stiffness of lava is described in terms of **viscosity**– lava that flows easily has low viscosity, and lava that is sticky and stiff has high viscosity.

In general, high-silica lava contains more gas than low-silica lava. When the gas forms into bubbles, viscosity increases further. Consider the pop analogy again. If you were to shake the bottle vigorously and then open it, the pop would come gushing out in a thick, frothy flow. In contrast, if you took care to not shake the bottle before opening it, you could pour out a thin stream of fluid.

Chemical Composition Affects the Thickness and Shape of Lava Flows

The thickness and shape of a lava flow depends on its viscosity. The greater the viscosity, the thicker the flow, and the shorter the distance it travels before solidifying. Highly viscous lava might not flow very far at all, and simply accumulate as a bulge, called a **lava dome**, in a volcano's crater. Figure 11.8 shows a dome formed from rhyolitic lava in the crater of Mt. St. Helens.

Figure 11.8 Lava dome in the crater of Mt. St. Helens. *Source: Terry Feuerborn (2011) CC BY-NC 2.0*

Less viscous rhyolitic lava can travel further, as with the thick flow in Figure 11.9 (right). The left of Figure 11.9 shows thin streams of freely flowing, low-silica, low-viscosity basaltic lava.

Figure 11.9 Lava flows. Left: A geologist collects a sample from a basaltic lava flow in Hawaii. Right: an andesitic lava flow from Kanaga Volcano in the Aleutian Islands. *Source: Left- U. S. Geological Survey (2014) Public Domain; Right- Michelle Combs, U. S. Geological Survey (2015) Public Domain*

Low-viscosity basaltic lava flows may travel extended distances if they move through conduits called **lava tubes**. These are tunnels within older solidified lava flows. Figure 11.10 (top) shows a view into a lava tube through a hole in the overlying rock, called a **skylight**. Figure 11.10 (bottom) shows the interior of a lava tube, with a person for scale. Lava tubes form naturally and readily because flowing mafic lava preferentially cools near its margins, forming solid **lava levées** that eventually close over the top of the flow. Lava within tubes can flow for 10s of km because the tubes insulate the lava from the atmosphere and slow the rate at which the lava cools. The Hawai'ian volcanoes are riddled with thousands of old, drained lava tubes, some as long as 50 km.

Figure 11.10 Lava tubes. Top: An opening in the roof of a lava tube (called a skylight) permitting a view of lava flowing through the tube (Pu‘u ‘Ō‘ō crater, Kīlauea). The opening is approximately 6 m across. Bottom: Inside a lava tube that channelled lava away from Mt. St. Helens in an eruption 1,895 years ago. *Sources: Top: U. S. Geological Survey (2016) Public Domain; Bottom: Thomas Shahan (2013) CC BY-NC 2.0*

Lava Structures

Pahoehoe

Lava flowing on the surface can take on different shapes as it cools. Basaltic lava with an unfragmented surface, like that in Figure 11.9 (right), is called **pahoehoe**. (pronounced pa-hoy-hoy). Pahoehoe can be smooth and billowy. It can also develop a wrinkled texture, called **ropy lava**, as shown in Figure 11.11. Ropy lava forms when the outermost layer of the lava cools and develops a skin (visible as a dark layer in Figure 11.11, left), but the skin is still hot and thin enough to be flexible. The skin is stiffer than the lava beneath it, and is dragged by flowing lava and folded up into wrinkles. Figure 11.11 (right) is a close-up view after a cut has been made to show the internal structure of a wrinkled lava flow. Notice the many holes, or **vesicles**, within the lava, formed when the lava solidified around gas bubbles.

Figure 11.11 Ropy lava from Hawaii. Left: Ropy texture forming as a thin surface layer of lava cools and is wrinkled by the motion of lava flowing beneath it. Right: Cross-section view of ropy lava. *Sources: Left: Z. T. Jackson (2005) CC BY NC-ND 2.0; Right: Fiddledydee (2011) CC BY-NC 2.0*

A'a and Blocky Lava

If the outer layer of the lava flow cannot accommodate the motion of lava beneath by deforming smoothly, the outer layer will break into fragments as lava moves beneath it. This could happen if the lava flow develops a thicker, more brittle outer layer, or if it moves faster. The result is a sharp and splintery rubble-like lava flow called **a'a** (pronounced like "lava" but without the l and v). Figure 11.12 (left) shows a close-up view of the advancing front of an a'a lava flow (the flow is moving toward the viewer). Figure 11.12 (right) shows an a'a lava flow viewed from the side. Compare the texture of the a'a flow with the texture of the lighter-grey pahoehoe lava in the foreground of the picture.

Figure 11.12 A'a lava flows. Left: Close-up view of a'a forming during an eruption of Pacaya Volcano in Guatemala. Field of view approximately 1 m across. Right: Rubbly reddish-brown a'a lava flow viewed from Chain of Craters Road, Hawai'i Volcanoes National Park. Pahoehoe is visible in lighter grey in the foreground. Sources: Photo of Hawai'ian a'a and pahoehoe: Roy Luck (2009) CC BY 2.0; Pacaya a'a: Greg Willis (2008) CC BY-SA 2.0 (labels added)

Higher viscosity andesitic lava flows also develop a fragmented surface, called **blocky lava**. This is visible in the toe of the andesitic lava flow from Figure 11.9 (right). The difference between a'a and the andesitic blocky lava is that the blocky lava has fragments with smoother surfaces and fewer vesicles.

Lava Pillows

When lava flows into water, the outside of the lava cools quickly, making a tube (Figure 11.13 (top left)). Blobs of lava develop at the end of the tube (Figure 11.13 (top right)), forming **pillows**. The bottom left of Figure 10.13 shows pillows covering the sea floor, and the bottom right shows the distinctive rounded shape of pillows in outcrop. Because pillows always form underwater, finding them in the rock record gives us information that the environment was underwater.

Figure 11.13 Pillow lavas. Top left: A tube of lava extruding underwater. Hot lava can be seen through cracks in the wall of the tube. The image is approximately 1 m across. (Pacific Ocean, near Fiji). Top right: The rounded end of a tube with cracks showing the lava within. (Pacific Ocean, near Fiji). Bottom left: sea floor near the Galápagos Islands covered with pillow lavas. Bottom right: A boulder made of 2.7 billion year old pillows derived from the Ely Greenstone in north-eastern Minnesota. *Sources: Top left- NSF and NOAA (2010) CC BY 2.0; Top right- NSF and NOAA (2010) CC BY 2.0; Bottom left- NOAA Okeanos Explorer Program, Galápagos Rift Expedition 2011 (2011) CC BY 2.0; Bottom right- James St. John (2015) CC BY 2.0*

Columnar Joints

When lava flows cool and solidify, they shrink. Long vertical cracks, or **joints**, form within the brittle rock to allow for the shrinkage. Viewed from above, the joints form polygons with 5, 6, or 7- sides, and angles of approximately 120º between sides (Figure 11.14). Figure 11.15 shows a side view of columnar joints in a basaltic lava flow in Iceland.

Left: **Figure 11.14** Columnar joints viewed from above, Giant's Causeway, Northern Ireland. *Source: Meg Stewart (2012) CC BY-SA 2.0*

Below: **Figure 11.15** Columnar joints in a basaltic lava flow, Svartifoss (Black Fall) Vatnajökull National Park, Iceland. Source: Ron Kroetz (2015) CC BY-ND 2.0

Pyroclastic Materials

The pop bottle analogy illustrates another key point about gas bubbles in fluid, which is that the bubbles can propel fluid. In the same way that shaking a pop bottle to make more bubbles will cause pop to gush out when the bottle is opened, gas bubbles can violently propel lava and other materials from a volcano, creating an explosive eruption.

Collectively, loose material thrown from a volcano is referred to as **tephra**. Individual fragments are referred to in general terms as **pyroclasts**, so sometimes tephra is also referred to as **pyroclastic debris**. Pyroclasts are classified according to size.

Volcanic Ash

Particles less than 2 mm in diameter are called **volcanic ash**. Volcanic ash consists of small mineral grains and glass. Figure 11.16 shows volcanic ash on three scales: in the upper left is ash from the 2010 eruption of Eyjafjallajökull in Iceland. The image was taken with a scanning electron microscope at approximately 1000 times magnification. In the upper right is ash from the 1980 eruption of Mt. St. Helens, collected in Yakima, Washington, about 137 km northeast of Mt. St. Helens. Individual particles are under 1 mm in size. Figure 11.16 (bottom) shows a village near Mt. Merapi in Indonesia dusted in ash after an eruption 2010.

Figure 11.16 Volcanic ash. Upper left: Ash from 2010 eruption of Eyjafjallajökull in Iceland, magnified approximately 1000x. Upper right- Ash from the 1980 eruption of Mt. St. Helens, collected at Yakima, Washington. Bottom: Indonesian village after the eruption of Mt. Merapi in 2010. *Sources: Upper left: Birgit Hartinger, AEC (2010) CC BY-NC-ND 2.0; Upper right: James St. John (2014) CC BY 2.0 (scale added); Bottom: AusAID/Jeong Park (2010) CC BY 2.0*

Lapilli

Fragments with dimensions between 2 mm and 64 mm are classified as **lapilli**. Figure 11.17 (upper left) shows lapilli at the ancient city of Pompeii, which was buried when Mt. Vesuvius erupted in 79 C.E. Figure 11.17 (lower left) is a form of lapilli called **Pele's tears**, named after the Hawai'ian diety Pele. Pele's tears form when droplets of lava cool quickly as they are flung through the air. Rapidly moving through the air may draw the Pele's tears out into long threads called **Pele's hair** (Figure 11.17, right). The dark masses in Figure 11.17 (right) within the Pele's hair are Pele's tears.

Figure 11.17 Lapilli are pyroclasts ranging between 2 mm and 64 mm in size. Upper left: lapilli from the site of the ancient city of Pompeii. Lower left: Pele's tears, a type of lapilli that forms when droplets of lava fly through the air. Right: Pele's hair, which form when Pele's tears are drawn out into thin threads as they fly. *Sources: Upper left: Pauline (2009) CC BY-NC-ND 2.0; Lower left: James St. John (2014) CC BY 2.0 (scale added); Right: James St. John (2009) CC BY 2.0 (scale added).*

Blocks and Bombs

Fragments larger than 64 mm are classified as blocks or bombs, depending on their origin. **Blocks** are solid fragments of the volcano that form when an explosive eruption shatters the pre-existing rocks. Figure 11.18 shows one of many blocks from an explosive eruption at the Halema'uma'u crater at Kīlauea Volcano in May of 1924. The block has a mass of approximately 7 tonnes and landed 1 km from the crater.

Figure 11.18 Volcanic block weighing approximately 7 tonnes thrown 1 km from the Halema'uma'u crater at Kīlauea Volcano on May 18, 1924. *Source: U. S. Geological Survey (1924) Public Domain*

Bombs form when lava is thrown from the volcano and cools as it travels through the air. Traveling through the air may cause the lava to take on a streamlined shape, as with the example in Figure 11.19.

Figure 11.19 Volcanic bomb with a streamlined shape. *Source: James St. John (2016) CC BY 2.0 (scale added)*

Effects of Gas on Lapilli and Bombs

The presence of gas in erupting lava can cause lapilli and bombs to take on distinctive forms as the lava freezes around the gas bubbles, giving the rocks a **vesicular** (hole-filled) texture. **Pumice** (Figure 11.20) forms from gas-filled felsic lava. Figure 11.20 (right), shows a magnified view of the sample on the left. The dark patches in the photograph are mineral crystals that formed in the magma chamber before the lava erupted. Pumice floats on water because some of the holes are completely enclosed, and air-filled.

Figure 11.20 Lapilli-sized pumice collected from the shores of Lake Atitlán in Guatemala by H. Herrmann. The lake is a flooded caldera, and is surrounded by active volcanoes. Right: Magnified view showing vesicular structure and amphibole crystals (dark patches). *Source: Karla Panchuk (2017) CC BY 4.0*

The mafic counterpart to pumice is **scoria** (Figure 11.21, left). Mafic lava can also form **reticulite** (Figure 11.21, right), a rare and fragile rock in which the walls surrounding the bubbles have all burst, leaving behind a delicate network of glass.

Figure 11.21 Vesicular mafic lapilli. Left: Scoria from Mount Fuji. Scoria is the mafic counterpart to pumice. Right: Reticulite is volcanic glass that forms when the walls separating gas bubbles pop. *Sources: Left- James St. John (2014) CC BY 2.0; Right- James St. John (2014) CC BY 4.0 (scales added)*

11.3 Types of Volcanoes

The products of volcanism that build volcanoes and leave lasting marks on the landscape include lava flows that vary in viscosity and gas content, and tephra ranging in size from less than a mm to blocks with masses of many tonnes. Individual volcanoes vary in the volcanic materials they produce, and this affects the size, shape, and structure of the volcano.

There are three types of volcanoes: **cinder cones** (also called **spatter cones**), **composite volcanoes** (also called **stratovolcanoes**), and **shield volcanoes**. Figure 11.22 illustrates the size and shape differences amongst these volcanoes.

Shield volcanoes, which get their name from their broad rounded shape, are the largest. Figure 11.22 shows the largest of all shield volcanoes- in fact, the largest of all volcanoes on Earth- Mauna Loa, which makes up a substantial part of the Island of Hawai'i and has a diameter of nearly 200 km. The summit of Mauna Loa is presently 4,169 m above sea level, but this represents only a small part of the volcano. It rises up from the ocean floor at a depth of approximately 5,000 m. Furthermore, the great mass of the volcano has caused it to sag downward into the mantle by an additional 8,000 m. In total, Mauna Loa is a 17,170 m thick accumulation of rock.

Figure 11.22 Comparison of volcano sizes and shapes. Broad, rounded shield volcanoes are the largest, followed by cone-shaped composite volcanoes. Straight-sided cinder cones are the smallest, and barely visible in the scale of the diagram. *Source: Karla Panchuk (2017) CC BY 4.0 modified after Steven Earle (2015) CC BY 4.0*

Kīlauea Volcano is also a shield volcano, albeit a much flatter one. Kīlauea Volcano rises only 18 m about the surrounding terrain, and is almost not visible in the scale of the diagram, however it still stretches over a distance of 125 km along the eastern side of the Island of Hawai'i.

Composite volcanoes are the next largest. Mt. St. Helens is shown on the left of Figure 11.22. It rises 1,356 m above the surrounding terrain in the Cascade Range of the western United States, and has a diameter of approximately 6 km. Composite volcanoes tend to be no more than 10 km in diameter. Unlike shield volcanoes, composite volcanoes have a distinctly conical shape, with sides that steepen toward the summit.

Cinder cones are the smallest, and almost too small to see next to a volcano like Mauna Loa. Eve Cone is a cinder cone on the flanks of Mt. Edziza in northwestern British Columbia. It rises 172 m above the landscape, and has a diameter of under 500 m. Cinder cones have straight sides, unlike upward-steepening composite volcanoes, or rounded shield volcanoes.

Volcano Structure

Shield Volcanoes

Shield volcanoes, like the Sierra Negra volcano in the Galápagos Islands (Figure 11.23, top), get their gentle hill-like shape because they are built of successive flows of low-viscosity basaltic lava (Figure 11.23, bottom). The low viscosity of the lava means that it can flow for long distances, resulting in the greater size of shield volcanoes compared to composite volcanoes or cinder cones.

Figure 11.23 Shield volcano. Top: The Sierra Negra volcano in the Galápagos Islands exhibits the low, rounded shape characteristic of shield volcanoes. Bottom: Diagram of a shield volcano island, showing the build up of basaltic lava flows. *Sources: Top- BRJ INC. (2012) CC BY-NC-ND 2.0; Bottom- Karla Panchuk (2017) CC BY 4.0*

Composite Volcanoes (Stratovolcanoes)

Composite volcanoes, like Cotopaxi in Figure 11.24 (top), consist of layers of lava alternating with layers of tephra (blocks, bombs, lapilli, and ash; Figure 11.24, bottom). The layers (strata) is where the alternative name, stratovolcano comes from. Cotopaxi displays the characteristic shape of composite volcanoes, which have slopes that get steeper near the top of the volcano. The change in the slope reflects the accumulation of tephra fragments near the volcano's vent. Composite volcanoes typically erupt higher viscosity andesitic and rhyolitic lavas, which do not travel as far from the vent as basaltic lavas do. This results in volcanoes of smaller diameter than shield volcanoes. A notable exception is Mt. Fuji in Japan, which erupts basaltic lava.

Figure 11.24 Composite volcano. Top: Cotopaxi in Ecuador exhibits the upward-steepening cone characteristic of composite volcanoes. Bottom: Diagram of a composite volcano showing alternating layers of lava and tephra. *Sources: Top- Photo by Simon Matzinger (2014) CC BY 2.0; Bottom: Karla Panchuk (2017) CC BY 4.0. See Appendix C for more attributions.*

From a geological perspective, composite volcanoes tend to form relatively quickly and do not last very long. If volcanic activity ceases, it might erode away within a few tens of thousands of years. This is largely because of the presence of pyroclastic eruptive material, which is not strong.

Cinder Cones (Spatter Cones)

Cinder cones, like Mt. Capulin in Figure 11.25, have straight sides and are typically less than 200 m high. Most are made up of fragments of scoria (vesicular rock from basaltic lava) that were expelled from the volcano as gas-rich magma erupted. Because cinder cones are made up almost exclusively of loose fragments, they have very little strength. They can be eroded away easily, and relatively quickly.

Figure 11.25 Cinder cone. These small, straight-sided volcanoes are made of volcanic fragments ejected when gas-rich basaltic lava erupts. *Sources: Karla Panchuk (2017) CC BY 4.0, with photograph by R. D. Miller, U. S. Geological Survey (1980) Public Domain. See Appendix C for more attributions.*

11.4 Types of Volcanic Eruptions

Volcanoes produce a variety of materials when they erupt. The characteristics of the eruptions themselves also vary from one volcano to the next, and sometimes from one eruption to the next for the same volcano. Eruptions are classified according how explosive they are, and the height of their **eruption column**- how high they blast material into the air.

Both the explosiveness of an eruption and the height of the eruption column are related in part to the composition of magma and the amount of gas it contains. In particular, magmas with more silica erupt more explosively. The higher the silica content, the greater the viscosity of the magma. This means more pressure can build up before the magma is forced out of the volcano. Magma with more silica also tends to contain more dissolved gas. The gas helps to propel magma out of the volcano, in the same way that the bubbles in a shaken bottle of pop cause the pop to foam out when the lid is removed.

There are four types of eruptions with properties determined mostly by the silica content of magma, and the amount of gas it contains. In order of increasing explosiveness, these are **Hawai'ian**, **Strombolian**, **Vulcanian**, and **Plinian** eruptions. Any composition of magma can have an explosive eruption if the magma suddenly encounters water. Hot magma contacting groundwater or seawater causes the water to flash to steam. Explosive eruptions driven by water are called **hydrovolcanic** (or **phreatic**) eruptions.

Hawai‘ian Eruptions

Hawai‘ian eruptions are named after the characteristic eruptions of volcanoes of the Hawai‘ian islands. Hawai‘ian eruptions are **effusive** (flowing) rather than explosive because they erupt low-viscosity basaltic

lava. Hawai'ian eruptions form shield volcanoes and can also take the form of **fissure eruptions**. Fissure eruptions occur when lava erupts from long cracks in the ground rather than from a central vent.

Figure 11.26 shows examples from two eruptions on of Hawai'i. In the upper left and right are images from the November 1959 eruption of Kīlauea Iki Crater. The upper left shows a fissure eruption and effusive flow of lava. Burning trees appear as bright spots toward the bottom of the photo. Figure 11.26 (right) shows a lava fountain reaching 425 m above Kīlauea Iki Crater. U. S. Geological Survey scientists reported that volcanic bombs up to 60 cm across smashed the guard rail and dented the asphalt on the road. Figure 11.26 (lower left) shows Hawaiian Volcano Observatory (HVO) scientists making a quick getaway, with lava fountains from Mauna Loa Volcano in the background.

The photographs of the Kīlauea Iki Crater and Mauna Loa Volcano eruptions make the point that while Hawai'ian eruptions are considered "gentle" eruptions, this is a relative term. "Gentle" eruptions range from lava flows that can be safely sampled by trained personnel, as in Figure 11.5, to lava fountains that soar hundreds of metres above the tree tops and rain large and dangerous rocks upon the surroundings.

Strombolian Eruptions

Strombolian eruptions, named for Mt. Stromboli in Italy, occur when basaltic lava has higher viscosity and higher gas content. The sticky lava is ejected in loud, violent, but short-lived spattery eruptions. Clumps of gas-rich lava thrown 10s to 100s of metres in the air accumulate as scoria in a pile around the vent, forming cinder cones. Figure 11.27 shows a strombolian eruption in the crater of Mt. Etna. A smaller cinder cone is forming around the vent as lava sputters out of it.

Figure 11.26 Hawai'ian eruptions. Top left: Fissure eruption at Kīlauea Iki Crater in November of 1959. Bottom left: Lava fountains from an eruption of Mauna Loa Volcano in 1984. Right: Lava fountain from Kīlauea Iki Crater eruption in November of 1959. *Sources: Top left- U. S. Geological Survey (1959) Public Domain; Bottom left: R. B. Moore, U. S. Geological Survey (1984) Public Domain; Right- U. S. Geological Survey (1959) Public Domain.*

Figure 11.27 Strombolian eruption of Mt. Etna. Sputtering lava forms a smaller cinder cone around a vent within the crater of Etna. *Source: Karla Panchuk (2017) CC BY-SA 4.0; Photograph; Robin Wylie (2012) CC BY 2.0. See Appendix C for more attributions.*

Vulcanian Eruptions

Vulcanian eruptions get their name from the volcanic Italian island of Vulcano, which itself takes the name of the Roman god of fire, Vulcan. In Roman mythology, Vulcan was the maker of armour and weaponry for the gods, and volcanic eruptions were attributed to him working in his forge.

Vulcanian eruptions are far more explosive than Strombolian eruptions, and can blast tephra and gas to a height of 5 to 10 km. The explosiveness is related to a build-up of pressure as the higher viscosity of intermediate silica content lava restricts the escape of gas. Vulcanian eruptions produce large quantities of ash in addition to blocks and bombs.

The Vulcanian eruption of Mt. Pelée on the island of Martinique in 1902 resulted in the first detailed documentation by geologists of a devastating phenomenon that is now referred to as a **pyroclastic flow** (Figure 11.28). Volcanic debris from the collapse of a lava dome on Mt. Pelée combined with hot gas to form a searing avalanche that raced down the mountain, over the city of St. Pierre, and into the harbour.

Figure 11.28 A series of photos taken by Alfred Lacroix during the eruption of Mt. Pelée on May 8, 1902 showing the development of the pyroclastic flow that destroyed the city of St. Pierre and nearly 30,000 inhabitants. *Source: Karla Panchuk (2017) CC BY 4.0. Photograph: Alfred Lacroix (1902) Public Domain. See Appendix C for more attributions*

The French geologist Alfred Lacroix described what he saw as a "nuée ardente," or thick fiery cloud. The following first-hand account was published in *Cosmopolitan Magazine* in July of 1902, attributed to Ellery S. Scott, a sailor on the steamship *Roraima*:

> *"In idle interest, I turned my glass toward Mont Pelee. It was at that very moment that the whole top of the mountain seemed blown into the air. The sound that followed was deafening. A great mass of flames, seemingly a mile in diameter, with twisting giant wreaths of smoke, rolled thousands of feet into the air, and then overbalanced and came rolling down the seamed and cracked sides of the mountain. Foot hills were overflowed by the onrushing mass. It was not mere flame and smoke. It was molten lava, giant blocks of stone and a hail of smaller stones, with a mass of scalding mud intermingled.*
>
> *For one brief moment I saw the city of St. Pierre before me. Then it was blotted out by the overwhelming flood. There was no time for the people to flee. They had not even time to pray.... I had called to Carpenter Benson to start the windlass, but before he could move, the "Roraima" rolled almost on her port beam-ends, and then as suddenly went to starboard. The funnel, masts and boats went by the board in an instant. The decks were swept clean. The hatches were staved in. The next instant a hail of fire and red-hot stones was upon the ship. Then came the scalding mud. The saloon was ablaze. The ship seemed doomed. Men were struck down all around me by flaming masses of lava. From bright sunlight the air became dense as midnight. The smoke that rolled down from the crater's mouth had blotted the sun from our vision."*

Scott's account vividly describes of the speed of the pyroclastic flow. In some cases, pyroclastic flows travel at speeds greater than 700 km/h. They are able to travel rapidly because they behave like a fluid, and can also ride on a cushion of hot gas. Scott says the city was "blotted out by a flood," yet the lower parts of buildings remained (Figure 11.29), and human remains were found in streets and homes where they had fallen. The ruins of St. Pierre look as though the top of the city were shaved off, and that is effectively what happened as the pyroclastic flow rushed across it, buoyed by gas.

Figure 11.29 Two stereographs of the ruins of St. Pierre, published in 1902. Stereographs are viewed with a stereoscope to make an image appear three dimensional. Top- "St. Pierre, 'the city of dead,' Mt. Pelee smoking, Martinique"; Bottom- "Overlooking the mud-filled Roxelane River bed, and ash-covered ruins, to Mont Pelée, St. Pierre, Martinique." *Source: Top- Boston Public Library (2013) CC BY 2.0; Bottom- Boston Public Library (2013) CC BY 2.0*

The vast majority of fatalities from the eruption were caused by the heat of pyroclastic flow. Examination of the ruins of St. Pierre revealed that glass had melted, but copper had not, putting the temperature at between 700 ºC and 1000 ºC (1292 ºF to 1832 ºF).

Plinian Eruptions

Plinian eruptions are explosive eruptions of intermediate to felsic lava, and can form eruptive columns up to 45 km high. The origin of the name is the eruption of Vesuvius in 79 CE, which buried the towns of Pompeii and Herculaneum. The Roman admiral Gaius Plinius Secundus, also known as Pliny the Elder, attempted a rescue mission when he saw the column of ash and debris above Vesuvius, but died of unknown causes without being able to reach Herculaneum.

A more recent Plinian eruption was that of Mt. Redoubt on April 21, 1990, shown in Figure 11.30. Pyroclastic flows resulted, as did **lahars**, landslides that formed when glaciers melted and turned volcanic ash into mud. The shape of the eruptive column, with parts of the column appearing to spread out in flat layers at different levels, reflects differences in atmospheric characteristics.

Figure 11.30 Plinian eruption of Mt. Redoubt in Alaska on April 21, 1990. *Source: Karla Panchuk (2017) CC BY 4.0. Photograph: R. Clucas, U. S. Geological Survey (1990) Public Domain. See Appendix C for more attributions.*

Hydrovolcanic (Phreatic) Eruptions

Hydrovolcanic eruptions can be far more explosive than Plinian eruptions. They occur when water in the form of groundwater, seawater, or even melting glacial ice or snow comes into contact with magma. Heat from the magma changes water suddenly to steam, which can expand to more than a thousand times the original volume of water. The sudden expansion results in an explosive force that can blast a volcano to pieces and create large amounts of volcanic ash.

In April of 2010, activity by the Icelandic volcano Eyjafjallajökull (Figure 11.31) melted the glacier above it, releasing large quantities of water and triggering a hydrovolcanic eruption. Ash rose in a plume 10 km high, and was blown westward and into the skies over Europe. Volcanic ash can damage or destroy aircraft engines, so the precaution was taken to prohibit air travel for a 5-day period. The enormous economic impact of stopping flights has led to numerous studies about the best way to deal with similar events with volcanic ash in the future.

Figure 11.31 Hydrovolcanic eruption of Eyjafjallajökull in April of 2010. Left- Eruptive column with volcanic lightning. Volcanic lightning is caused by the static electricity generated by volcanic ash particles rubbing together. Right- Another view of the ash cloud, with westward winds carrying ash toward Europe where it would disrupt air traffic. *Source: Karla Panchuk (2017) CC BY-SA 4.0. Left photograph: Terje Sørgjerd (2010) CC BY-SA 3.0; Right photograph: Henrik Thorburn (2010) CC BY 3.0. See Appendix C for more attributions.*

11.5 Plate Tectonics and Volcanism

Thus far volcanoes have been discussed in terms of the kinds of volcanic mountains they form, the materials they produce, and the style of eruption they have. All of these characteristics can be tied together into a big picture by considering the plate tectonic settings in which magma forms (Figure 11.32). The vast majority of volcanoes are present along plate tectonic boundaries.

Figure 11.32 Plate tectonic settings of volcanism. Volcanoes along subduction zones are the result of flux melting (lowering the melting point by adding water). Decompression melting produces volcanoes along divergent margins (ocean spreading centres and continental rift zones), as well as above mantle plumes. Contact between hot mafic partial melts and felsic rocks can trigger partial melting of the felsic rocks (melting from conduction). *Source: Karla Panchuk (2017) CC BY 4.0. Modified after Steven Earle (2015) CC BY 4.0 and U. S. Geological Survey (1999) Public Domain.*

There are four main scenarios to consider:

- Divergent boundaries where melting is triggered by decompression
- Subduction zones (ocean-ocean and ocean-continent convergent boundaries) where flux melting occurs as water is released from subducting ocean crust
- **Hot spots** where plumes of hot mantle material rise up, then melt as a result of decompression.
- Melting by conduction when magma transfers heat to rocks having a lower melting temperature.

Decompression Causes Volcanism Along Spreading Centres and Rift Zones

At an ocean spreading ridge (centre of Figure 11.32), convection moves hot mantle rock slowly upward at rates of cm per year. At roughly 60 km below the surface, the mantle rocks have decompressed is enough to permit partial melting of approximately 10% of the ultramafic rock. Mafic magma is produced, and it moves up toward the surface. Magma fills vertical fractures produced by the spreading and spills out onto the sea floor making pillow lavas and lava flows. Spreading-ridge volcanism is taking place approximately 200 km offshore from the west coast of Vancouver Island.

In continental rift zones where continental crust is thinning (far right in Figure 11.32), a similar decompression process occurs, triggering partial melting of ultramafic mantle rocks. However, if the continental crust above the region where melting occurs has a lower melting temperature than the mafic melt that is produced, the continental crust will also melt.

Continental rift zones can have a range of volcano types. If mafic magma erupts, shield volcanoes, broad lava flows, and cinder cones result. However, if rocks of other compositions are melted and added in, or the mafic magma undergoes fractional crystallization before erupting, then composite volcanoes will also form.

Water Causes Partial Melting Along Subduction Zones

At an ocean-continent convergent boundary (Figure 11.32, right) or ocean-ocean convergent boundary (Figure 11.32, left), oceanic crust is pushed down into the mantle. Although temperatures are high, the slab is kept from melting by high pressures. However, under these conditions minerals in the slab release water from within their crystal structures. The water lowers the melting point of rock above the slab, and partial melting is triggered within the mantle. Mafic magma rises through the mantle to the base of the crust. There it contributes to partial melting of crustal rock, and more felsic material is added to the magma. The magma, now intermediate in composition, continues to rise and assimilate crustal material. In the upper part of the crust, it accumulates into plutons. Over time, fractional crystallization of magma within the pluton can make it even more silica-rich. From time to time, the magma from the plutons rises toward surface, leading to volcanic eruptions.

Composite volcanoes with Vulcanian or Plinian eruption styles are characteristic of the volcanic arcs that form in subduction zones, although in the Trans-Mexico Volcanic Belt, Strombolian eruptions produce short-lived cinder cones. Where two margins of oceanic crust collide, the volcanic arc will be a chain of volcanic islands. Where continental and oceanic crust collide, there will be a volcanic arc on the continental crust.

Mt. St. Helens: A Composite Volcano in the Cascades Range Continental Volcanic Arc

On May 18, 1980 at 8:32 a.m., a M5.1 earthquake shook Mt. St. Helens, and marked the start of a 9-hour Plinian eruption (Figure 11.33) with a 24 km high eruption column and multiple pyroclastic flows. By the time the eruption was over, a large part of the volcano had been blasted away.

Figure 11.33 Eruption of composite subduction-zone volcano Mt. St. Helens on May 18, 1980. Top- Plinian eruption column. Bottom left- Mt. St. Helens before the eruption. Bottom right- The remains of Mt. St. Helens after the eruption. *Sources: Top- Karla Panchuk (2017) CC BY 4.0; Photograph by NOAA (1980) Public Domain view source. Bottom left- R. Hoblitt, U. S. Geological Survey, Cascades Volcano Observatory (1979) Public Domain (label added); Bottom right- Steven Earle (2015) CC BY 4.0 (label added). See Appendix C for more attributions.*

The explosive eruption was driven by gas-rich rhyolitic magma, however not all of Mt. St. Helens' eruptions have been of felsic or intermediate material. The lava tube in Figure 11.10 (bottom) is from a time when Mt. St. Helens erupted basaltic lava. Data from the iMUSH (Imaging Magma Under St. Helens) project show that a magma chamber is present beneath Mt. St. Helens at between 5 and 14 km depth, but that a much larger magma chamber is present below it, which extends down to the mantle (Figure 11.34). Earthquakes in the 24 hours after the 1980 eruption suggested movement of magma within the smaller chamber (yellow arrows in Figure 11.34), but earthquakes from 1980 to 2005 indicate movement of magma within the deeper chamber as well (black arrows).

The complex history of Mt. St. Helens could reflect changes in the composition of magma within the small chamber over time, as fractionation proceeds, and the magma becomes more silica rich. However, movement of more mafic magma from the larger chamber could also contribute to eruptions with different chemical compositions. The larger magma chamber may be connected to a chamber feeding the nearby Indian Heaven Volcanic Field, which contains shield volcanoes and cinder cones, and for which basalt makes up 80% of erupted materials.

Mt. St. Helens, continued

Figure 11.34 Magma chambers beneath Mt. St. Helens and Indian Heaven Volcanic Field, sketched from iMUSH (Imaging Magma Under St. Helens) project results. In a 24 hour period after the May 18, 1980 eruption, earthquakes in and around the smaller magma chamber suggested migration of magma (yellow arrows). Earthquakes recorded between 1980 and 2005 suggest migration of magma within a larger chamber that extends to the mantle (black arrows). The larger magma chamber might feed another smaller chamber beneath the Indian Heaven Volcanic Field. *Source: Karla Panchuk (2017) CC BY 4.0, based on Kiser et al. (2016), Figure 4B.*

Mantle Plumes Can Cause Volcanism Away from Plate Boundaries

Mantle plumes are rising columns of hot solid rock. The column may be kilometres to 10s of kilometres across, but near the surface it spreads out to create a mushroom-like head that is 10s to over 100 kilometres across. Mantle plumes are different from the convection that normally occurs beneath ocean spreading centres: plumes rise approximately 10 times faster than mantle convection normally occurs, and may originate deep in the mantle, possibly just above the core-mantle boundary.

When the mantle plume rises to the base of the lithosphere, the pressure is low enough to permit partial melting of the plume material, producing mafic magma. Heat carried by the mantle plume may also melt rock adjacent to the plume. The magma rises and feeds **hotspot volcanoes**. The lithospheric plate above the mantle plume is moving across the plume, so a chain of hotspot volcanoes can result as existing hotspot volcanoes are slowly moved away from the mantle plume, and new volcanoes form in the lithosphere.

Many shield volcanoes are associated with mantle plumes, including those that make up the Hawai'ian islands. All of the Hawai'ian volcanoes are related to the mantle plume that currently lies beneath Mauna Loa, Kilauea, and Lō‘ihi (Figure 11.35, top). There is evidence of crustal magma chambers beneath all three active Hawai'ian volcanoes. At Kīlauea, the magma chamber appears to be several kilometres in diameter, and is situated between 8 km and 11 km below surface (Lin et al., 2014). In this area, the Pacific Plate is moving northwest at a rate of about 7 cm/year. This means that the earlier formed — and now extinct — volcanoes have now moved well away from the mantle plume. The hotspot has in fact been present for at least 85 million years (Regelous et al., 2003), as evidenced by the long chain of eroded and submerged mountains stretching to the Aleutian Trench (Figure 11.35, bottom).

Figure 11.35 Hawai'ian hotspot volcanoes and volcanic chain. Top- A mantle plume beneath Hawai'i supplies magma to Mauna Loa Volcano, Kīlauea Volcano, and Lō'ihi Seamount. Volcanoes to the northwest are no longer active because they have moved away from the plume. Bottom- Bathymetric (depth) map showing the chain of islands stretching toward the Aleutian Trench, and marking the progress of the Pacific Plate over the mantle plume. *Source: Top- J. E. Robinson, U. S. Geological Survey (2006) Public Domain; Bottom- National Geophysical Data Center/ U. S. Geological Survey (2006) Public Domain (labels added).*

Kīlauea Volcano is approximately 300 ka old, while neighbouring Mauna Loa Volcano is over 700 ka and Mauna Kea Volcano is over 1 Ma. If volcanism continues above the Hawaii mantle plume in the same manner that it has for the past 85 Ma, it is likely that Kīlauea Volcano will continue to erupt for at least another 500,000 years. By that time, its neighbour, Lō'ihi Seamount, will have emerged from the sea floor, and its other neighbours, Mauna Loa and Mauna Kea, will have become significantly eroded, like their cousins, the islands to the northwest.

Large Igneous Provinces (LIPs)

While the Hawaii mantle plume has produced a relatively low volume of magma for approximately 85 Ma, other mantle plumes are less consistent, and some generate massive volumes of magma over relatively short time periods. Although their origin is still controversial, it is thought that the volcanism leading to **large igneous provinces** (LIPs) is related to very high volume but relatively short duration bursts of magma from mantle plumes. An example of an LIP is the Columbia River Basalt Group, which extends across Washington, Oregon, and Idaho in the United States (Figure 11.36). This volcanism, which covered an area of about 160,000 km^2 with basaltic rock up to several hundred metres thick, took place between 17 and 14 Ma.

Figure 11.36 Part of the Columbia River Basalt Group at Frenchman Coulee, eastern Washington, United States. All of the flows visible here have formed large (up to two metres in diameter) columnar basalts, a result of relatively slow cooling of flows that are tens of m thick. The inset map shows the approximate extent of the 17 to 14 Ma Columbia River Basalts, with the location of the photo shown as a star. *Source: Steven Earle (2015) CC BY 4.0*

The mantle plume that is assumed to be responsible for the Columbia River LIP is now situated beneath the Yellowstone area, where it leads to felsic volcanism. Over the past 2 Ma, three very large explosive eruptions at Yellowstone have yielded approximately 900 km^3 of felsic magma. This is approximately 900 times the volume of the 1980 eruption of Mt. St. Helens, but only 5% of the volume of mafic magma in the Columbia River LIP.

Most other LIP eruptions are much bigger. The Siberian Traps (also basalt), which erupted at the end of the Permian period at 251 Ma, are estimated to have produced approximately 40 times as much lava as the Columbia River LIP. The largest known LIP is the Ontong Java Plateau, located in the southwest Pacific Ocean. It formed at around 122 Ma, and presently covers 1,500,000 km^2 and has a volume of 5,000,000 km^3. But this is only a small fraction of its original size. The majority of it has been subducted, and it may have been split into pieces that have been classified as separate LIPs.

Kimberlites

Kimberlite pipes are carrot-shaped cones of ultramafic rock. They form from the explosive eruption of mantle plumes originating at depths of 150 to 450 km in the mantle. The plume makes its way to the surface quickly (over hours to days), having little interaction with the surrounding rocks, and thus preserving a sample of the ultramafic mantle. As the plume nears the surface, a build-up of gas causes it to pick up speed, and by the time it reaches the surface it may be travelling faster than the speed of sound. The explosiveness of kimberlite eruptions means that they do not form volcanic mountains on the surface, but leave circular holes in the ground.

Kimberlite eruptions that originate at depths greater than 200 km beneath old, thick, continental crust travel through the region of the mantle where diamond is stable. In some cases, such as in Saskatchewan and the

Northwest Territories, kimberlites bring diamond-bearing material to the surface. All of Earth's diamond deposits are thought to have originated in this way.

Diamond mines in kimberlites, such as the Ekati Mine in the Northwest Territories, are easy to spot by the characteristic circular hole that develops as miners excavate the cone-shaped structure (Figure 11.37). The kimberlites at Ekati erupted between 45 and 60 Ma. Many kimberlites are older, and some much older. There have been no kimberlite eruptions in historic times. The youngest known kimberlites are in the Igwisi Hills in Tanzania and are only about 10,000 years old. The next youngest date to approximately 30 Ma.

Figure 11.37 The Ekati diamond mine in the Northwest Territories, part of the Lac de Gras kimberlite field. *Source: Karla Panchuk (2017) CC BY-SA 4.0; Photograph by J. Pineau (2010) CC BY-SA 3.0. See Appendix C for more attributions.*

11.6 Volcanic Hazards

The basaltic lava flows produced by volcanoes on the island of Hawai'i are responsible for extensive damage to homes, infrastructure, and habitats. Figure 11.38 shows lava flows (in black) from the Pu‘u ‘Ō‘ō crater of Kīlauea Volcano. The lava flow destroyed the house, and is encroaching on the transfer station. Smoke in the background marks locations where additional flows have broken out and are burning vegetation.

Figure 11.38 Lava flow from Kīlauea's Pu‘u ‘Ō‘ō crater. Lava (in black) has destroyed a house and threatens a transfer station. *Source: U. S. Geological Survey (2014) Public Domain*

In spite of the damage that lava flows can cause, they are not the volcanic hazard with the greatest impact on lives and infrastructure. Even the relatively free-flowing Hawai'ian basaltic lava moves slowly enough that it can be escaped on foot. Far more dangerous hazards are related to gases and volcanic debris. However, the largest impact and the greatest suffering are caused not by the immediate effects of volcanic eruptions, but by large-scale changes to climate and environments caused by volcanism. Indirect effects resulting in respiratory distress, toxicity, famine, and habitat destruction have accounted for approximately 8 million deaths during historical times, while direct effects have accounted for fewer than 200,000, or 2.5% of the total.

Volcanic Gas and Tephra Emissions

Large volumes of rock and gases are emitted during major Plinian eruptions at composite volcanoes, and a large volume of gas is released during some very high-volume effusive eruptions. Gases and fine particles of volcanic ash can cause respiratory distress and poisoning, and ash poses a risk for aircraft.

Most of the tephra from large explosive eruptions ascends high into the atmosphere, and some of it is distributed around Earth by high-altitude winds. The larger components (larger than 0.1 mm) fall closer to the volcano, and the accumulation of tephra from large eruptions can cause serious damage and casualties. When the large eruption of Mt. Pinatubo in the Philippines occurred in 1991, tens of centimetres of ash accumulated in fields and on rooftops in the surrounding populated region. Heavy typhoon rains hit the island at the same time and added to the weight of the tephra. The weight was too much for roofs to bear, and thousands of structures collapsed, causing at least 300 of the 700 deaths attributed to the eruption.

One of the long-term effects of adding volcanic particles and gases to the atmosphere is cooling. Over an eight-month period in 1783 and 1784, a massive effusive eruption took place at the Laki volcano in Iceland. Although there was relatively little volcanic ash involved, a massive amount of sulphur dioxide was released into the atmosphere, along with a significant volume of hydrofluoric acid (HF). The sulphur dioxide combined with water to make sulphate **aerosols**, which block incoming solar energy. The accumulation of sulphate aerosols over that 8 months led to dramatic cooling in the northern hemisphere. There were serious crop failures in Europe and North America, and a total of 6 million people are estimated to have died from famine and respiratory complications. In Iceland, poisoning from the HF resulted in the death of 80% of sheep, and 50% of cattle. The ensuing famine, along with HF poisoning, resulted in more than 10,000 human deaths, about 25% of the population.

Pyroclastic Flows

In a typical explosive eruption at a composite volcano, the tephra and gases are ejected with explosive force and sent high up into the atmosphere. As the eruption proceeds, and the amount of gas in the rising magma starts to decrease, and less gas is supplied to the eruption column. Parts of the column will become denser

than air, leading the column to collapse and flow downward along the flanks of the volcano (Figure 11.39), picking up speed as it cools.

Figure 11.39 The Plinian eruption of Mt. Mayon, Philippines in 1984. Although most of the eruption column is ascending into the atmosphere, pyroclastic flows are traveling down the sides of the volcano in several places. Warnings were issued in time to evacuate 73,000 people. *Source: C. G. Newhall, U. S. Geological Survey (1984) Public Domain*

Pyroclastic flows can travel over water, in some cases for many kilometres. In 1902 the pyroclastic flow from the eruption of Mt. Pelée traveled out into the harbour and destroyed several wooden ships anchored there. The pyroclastic flow from the 1883 eruption of Krakatau traveled 80 km across the Sunda Straits and claimed victims on the southwest coast of Sumatra. It also triggered a tsunami.

One of the most famous pyroclastic flows occurred when Mt. Vesuvius erupted in 79 CE. It buried the cities of Pompeii and Herculaneum, killing an estimated 18,000 people.

Lahar

A **lahar** is any **mudflow** or **debris flow** that is related to a volcano (Figure 11.40). Most are caused by melting snow and ice during an eruption, as was the case with the lahar that destroyed the Colombian town of Armero in 1985 when the volcano Nevado del Ruiz caused the ice dam on a glacial lake to fail. The resulting lahar killed 23,000 in Armero, about 50 km from the volcano.

Figure 11.40 Mud left behind from the lahar after the May 18, 1980 eruption of Mt. St. Helens. The lahar carried the boulder to its present location. *Source: L. Topinka, U. S. Geological Survey (1980) Public Domain*

Lahars can also happen when there is no volcanic eruption, because composite volcanoes tend to be weak and easily eroded. In October 1998, category 5 hurricane Mitch slammed into the coast of Central America. Damage was extensive and 19,000 people died. Fatalities were largely because of mudflows and debris flows triggered by intense rainfall — some regions received almost 2 m of rain over a few days.

At Casita Volcano in Nicaragua, the heavy rains weakened rock and volcanic debris on the upper slopes, resulting in a debris flow that rapidly built in volume as it raced down the steep slope. It struck the towns of El Porvenir and Rolando Rodriguez killing more than 2,000 people. El Porvenir and Rolando Rodriguez were new towns that had been built without planning approval in an area that was known to be at risk of lahars.

Sector Collapse and Debris Avalanche

In the context of volcanoes, **sector collapse** or **flank collapse** is the catastrophic failure of a significant part of an existing volcano, creating a large debris avalanche. This hazard was first recognized with the failure of the north side of Mt. St. Helens immediately prior to the large eruption on May 18, 1980.

In the weeks before the eruption, a large bulge had formed on the side of the volcano (Figure 11.41) as magma moved from depth into a magma chamber within the mountain itself. Early on the morning of May 18, a moderate earthquake struck and destabilized the bulge, leading to Earth's largest observed landslide in historical times. The failure of this part of the volcano exposed the underlying magma chamber, causing it to explode sideways. This in turn exposed the conduit leading to the magma chamber below, resulting in a Plinian eruption lasting nine hours.

Figure 11.41 Bulge forming on the north side of Mt. St. Helens, April 27 1980. *Source: P. Lipman, U. S. Geological Survey (1980) Public Domain*

11.7 Monitoring Volcanoes and Predicting Eruptions

In 2005 U. S. Geological Survey geologist Chris Newhall made a list of the six most important signs of an imminent volcanic eruption. They are:

1. *Gas leaks* — the release of gases (mostly H_2O, CO_2, and SO_2) from the magma into the atmosphere through cracks in the overlying rock
2. *Bulging* — the deformation of part of the volcano, indicating that a magma chamber at depth is swelling or becoming more pressurized
3. *Seismicity* — many (hundreds to thousands) of small earthquakes, indicating that magma is on the move. The quakes may be the result of the magma forcing the surrounding rocks to crack, or a harmonic vibration that is evidence of magmatic fluids moving underground.
4. *Seismicity ceases* — a sudden decrease in the rate of earthquake activity. This may indicate that magma has stalled, and that\ something is about to give way

5. *Big bump* — a pronounced bulge on the side of the volcano (like the one at Mt. St. Helens in 1980), which may indicate that magma has moved close to surface
6. *Steam* — steam eruptions (**phreatic eruptions**) that happen when magma near the surface heats groundwater to the boiling point. The water eventually explodes, sending fragments of the overlying rock far into the air.

With these signs in mind, it is possible to determine the necessary equipment to have and actions to take to monitor a volcano and predict when it might erupt. we can make a list of the equipment we should have and the actions we can take to monitor a volcano and predict when it might erupt.

Assessing Seismicity

The simplest and cheapest way to monitor a volcano is with **seismometers**, instruments that detect vibration. In an area with several volcanoes that have the potential to erupt (e.g., the Squamish-Pemberton area), a few well-placed seismometers can provide an early warning that something is changing beneath one of the volcanoes. There are currently enough seismometers in the Lower Mainland and on Vancouver Island to provide this information. You can view a map of Canadian National Seismograph Network here.

If there is seismic evidence that a volcano is coming to life, more seismometers should be placed in locations within a few tens of kilometres of the source of the activity (Figure 11.42). This will allow geologists to determine the exact location and depth of the seismic activity so that they can see where the magma is moving.

Figure 11.42 A seismometer installed in 2007 in the vicinity of the Nazco Cone, BC. *Source: Cathie Hickson (n.d.) used with permission.*

Detecting Gases

Water vapour quickly turns into clouds of liquid water droplets and is relatively easy to detect just by looking, but CO_2 and SO_2 are not as obvious. It's important to be able to monitor changes in the composition of volcanic gases, and we need instruments to do that. Some can be monitored from a distance (from the ground or even from the air) using infrared devices, but to obtain more accurate data, we need to sample the air and do chemical analysis. This can be achieved with instruments placed on the ground close to the source of the gases, or by collecting samples (Figure 11.43) and analyzing them in a lab.

Figure 11.43 A geologist collects a gas sample from Sherman Crater, Mt. Baker, Washington. Gas is drawn through a titanium tube inserted in a fumarole, and collected in a glass vacuum flask. *Source: D. Tucker, U. S. Geological Survey (2006) Public Domain*

Measuring Deformation

There are two main ways to measure ground deformation at a volcano. One is known as a **tiltmeter**, which is a sensitive three-directional level that can sense small changes in the tilt of the ground at a specific location. Another is through the use of GPS (global positioning system) technology (Figure 11.27). GPS is more effective than a tiltmeter because it provides information on how far the ground has actually moved — east-west, north-south, and up-down.

Figure 11.44 A GPS unit installed at Hualālai Volcano, Hawaii. The dish-shaped antenna on the right is the GPS receiver. The antenna on the left is for communication with a base station. *Source: U. S. Geological Survey (n.d.) Public Domain*

Putting It All Together

By combining information from these types of sources, along with careful observations made on the ground and from the air, and a thorough knowledge of how volcanoes work, geologists can get a good idea of the potential for a volcano to erupt in the near future (months to weeks, but not days). They can then make recommendations to authorities about the need for evacuations and restricting transportation corridors.

Our ability to predict volcanic eruptions has increased dramatically in recent decades because of advances in our understanding of how volcanoes behave and in monitoring technology. Providing that careful work is done, there is no longer a large risk of surprise eruptions, and providing that public warnings are issued and heeded, it is less and less likely that thousands will die from sector collapse, pyroclastic flows, ash falls, or lahars. Indirect hazards are still very real, however, and we can expect the next eruption like the one at Laki in 1783 to take an even greater toll than it did then, especially since there are now roughly eight times as many people on Earth.

11.8 Volcanoes in Canada

Canada's volcanically active regions are located in British Columbia and the Yukon Territory (Figure 11.45). At least 49 eruptions have occurred within these regions in the last 10,000 years. There are five volcanic regions associated with three types of plate tectonic settings: a subduction zone, a mantle plume, and a continental rift zone.

Subduction Zone Volcanism: Wrangell and Garibaldi Volcanic Belts

The Wrangell Volcanic Belt is the result of subduction beneath the North American Plate. Volcanoes in the Canadian part of the Wrangell Volcanic Belt erupted between 17.8 and 10.4 million years ago. They were fed by lava that seeped up along a leaky transform fault.

Left **Figure 11.45** Canada's volcanic regions are located in British Columbia and the Yukon Territory. Volcanism is associated with three tectonic settings: the subduction zone along the west coast (Garibaldi Volcanic Belt, Wrangell Volcanic Belt), a continental rift zone (Wells Gray-Clearwater Volcanic Field, Stikine Volcanic Belt), and a mantle plume (Anahim Volcanic Belt). *Source: Volcanoes Canada, Canadian Hazards Information Service, Natural Resources Canada (n.d.). See Appendix C for copyright information.*

Below **Figure 11.46** Volcanic centres within the Garibaldi Volcanic Belt. The most recent eruption was 2,350 years ago at Mt. Meager. *Source: Sémhur (2007) CC BY-SA 4.0.*

Southwestern British Columbia is at the northern end of the Juan de Fuca subduction zone, and part of the Cascade Volcanic Arc that extends south through Washington and Oregon. The Canadian part of the Cascade Arc has had a lower rate and volume of volcanism than U. S. portions. One reason is that the northern part of the Juan de Fuca Plate is subducting more slowly than the rest of the plate, or else has stalled.

The Garibaldi Volcanic Belt has several **volcanic centres** , or regions where volcanism has caused multiple volcanoes to develop (Figure 11.46).

The most recent volcanic activity in this area was 2,350 years ago at Mt. Meager. An explosive eruption similar in magnitude to that of Mt. St. Helens in 1980 spread ash as far east as Alberta. There was also significant volcanic activity at Mt. Price and Mt. Garibaldi approximately 10,000 years ago as glacial ice receded. In both cases, lava and tephra built up against glacial ice. The western side of Mt. Garibaldi failed by sector collapse when the ice melted, leaving rocks unsupported. Eruption beneath glacial ice resulted in the formation of a **tuya**- a steep-sided, flat-topped volcano- called The Table near Mt. Garibaldi (Figure 11.47).

Figure 11.47 The Table, a tuya near Mt. Garibaldi. Tuyas form when volcanoes erupt beneath ice, and their shape is determined by rapid cooling beneath the ice sheet. *Source: Andre Charland (2004) CC BY 2.0*

Mantle Plume Volcanism: Anahim Volcanic Belt

The chain of volcanic complexes and cones extending from Milbanke Sound to Nazko Cone is interpreted as being related to a mantle plume currently situated close to the Nazko Cone, just west of Quesnel (Figure 11.48). The North American Plate is moving in a westerly direction at about 2 cm per year with respect to this plume, and the series of now partly eroded shield volcanoes between Nazco and the coast is interpreted to have been formed by the plume as the continent moved over it.

The Rainbow Range, which formed at approximately 8 Ma, is the largest of these older volcanoes. It has a diameter of about 30 km and an elevation of 2,495 m (Figure 11.49). The name "Rainbow" refers to the bright colours displayed by some of the volcanic rocks as they weather.

Top- **Figure 11.48** Anahim Volcanic Belt, the result of a mantle plume beneath the North American Plate. *Source: Sémhur (2007) CC BY-SA 4.0*

Left- **Figure 11.49** Tsitsutl, the "painted mountain" within the Rainbow Range of the Anahim Volcanic Belt. The vibrant colours of the Rainbow Range are the result of chemical weathering. *Source: Drew Brayshaw (2015) CC BY-NC 2.0*

Rift-Related Volcanism: Wells Gray-Clearwater Volcanic Field and Stikine Volcanic Belt

While British Columbia is not about to split into pieces, two areas of volcanism are related to rifting, or at least to stretching-related fractures that might extend through the crust. These are the Wells Gray-Clearwater volcanic field southeast of Quesnel (Figure 11.50), and the Stikine Volcanic Belt (also called the Northern Cordillera Volcanic Province), which ranges across the northwestern corner of the province.

The Stikine Volcanic Belt includes Canada's most recent volcanic eruption, a cinder cone and mafic lava flow that formed around 250 years ago at the Tseax River Cone in the Nass River area north of Terrace. According to Nisga'a oral history, lava overran a village on the Nass River, and 2,000 people were lost. The region is now part of the Anhluut'ukwsim Laxmihl Angwinga'asanskwhl Nisga'a (Nisga'a Memorial Lava Bed Park).

The Mount Edziza Volcanic Field near the Stikine River is a large area of lava flows, sulphurous ridges, and cinder cones. The most recent eruption in this area was about 1,000 years ago. While most of the other volcanism in the Edziza region is mafic and involves lava flows and cinder cones, Mt. Edziza itself (Figure 11.51) is a composite volcano with rock compositions ranging from rhyolite to basalt. A possible explanation for the presence of composite volcanism in an area dominated by mafic flows and cinder cones is that there is a magma chamber beneath this area, within which magma differentiation is taking place.

Figure 11.50 Wells Gray-Clearwater Volcanic Field is the result of extension in the crust.
Source: Sémhur (2007) CC BY-SA 4.0

Figure 11.51 Mount Edziza, in the Stikine Volcanic Belt, BC, with Eve Cone in the foreground.
Source: NASS5518 (2008) CC BY 2.0

Chapter 11 Summary

The topics covered in this chapter can be summarized as follows:

11.1 What Is A Volcano?

Volcanoes are places where molten rock escapes to Earth's surface. Some volcanoes are cone-shaped or hill-shaped mountains, and some eruptions happen along fissures. Eruptions are fed by a magma chamber beneath the volcano. Sometimes a volcano collapses into empty space in the magma chamber beneath, forming a caldera.

11.2 Materials Produced by Volcanic Eruptions

Volcanoes produce gas, lava flows, and debris called tephra. The characteristics of a lava flows depend on whether the lava is thin and runny (mafic with low gas content) or thick and sticky (felsic with high gas content). Tephra is classified according to size. Ash is less than 2 mm in diameter, lapilli is between 2 mm and 64 mm, and blocks and bombs are larger than 64 mm.

11.3 Types of Volcanoes

Cinder cones are relatively small straight-sided volcanoes that are composed mostly of mafic rock fragments. Composite volcanoes consist of alternating layers of lava flows and tephra. The tend to be intermediate to felsic in composition, and get steeper toward the top. Shield volcanoes are broad, low, hill-like volcanoes that form from layers of low-viscosity mafic lava.

11.4 Types of Volcanic Eruptions

Volcanic eruptions can be classified according to how explosive they are, and how high into the atmosphere they blast material. Hawai'ian eruptions are relatively gentle effusive eruptions of low-viscosity mafic lava, and form shield volcanoes. Strombolian eruptions are more vigorous eruptions of mafic tephra. They blast material hundreds of metres into the air. The tephra falls out of the atmosphere to form a cinder cone. Vulcanian eruptions are explosive eruptions of intermediate composition lava, producing pyroclastic flows and eruptive columns from 5 to 10 km high. Plinian eruptions are highly explosive eruptions of felsic lava, and can produce eruption columns up to 45 km high. Both Vulcanian and Plinian eruptions are associated with composite volcanoes. Hydrovolcanic eruptions are the explosive result of magma or lava interacting with water, and rapidly changing the water to steam.

11.5 Plate Tectonics and Volcanism

Volcanism is closely related to plate tectonics. Most volcanoes are associated with convergent plate boundaries (at subduction zones), but a great deal of volcanic activity also occurs at divergent boundaries and areas of continental rifting. At convergent boundaries magma is formed where water from a subducting plate acts as a flux to lower the melting temperature of the adjacent mantle rock. At divergent boundaries magma forms because of decompression melting. Decompression melting also takes place within a mantle plume.

11.6 Volcanic Hazards

Most direct volcanic hazards are related to volcanoes that erupt explosively, especially composite volcanoes. Pyroclastic flows, some as hot as 1000 ˚C, can move at hundreds of km/h and will kill anything in the way. Lahars, volcano-related mudflows, can be large enough to destroy entire towns. Lava flows are also destructive, but tend to move slowly enough to permit people to get to safety. Indirect hazards claim far more lives than direct hazards, and include famine related to volcanically-induced climate cooling.

11.7 Monitoring Volcanoes and Predicting Eruptions

Clues that a volcanic eruption might soon occur include earthquakes, a change in the type and amount of gases released, and changes in the shape of the volcano as magma moves within it. Volcanoes are monitored using seismometers to detect earthquakes, volcanic gases are sampled and analyzed, and instruments are used to detect deformation of the volcano. These tools make it possible to assess the hazard posed by a given volcano, and the risk of eruption.

11.8 Volcanoes in Canada

British Columbia and the Yukon Territory include examples of volcanoes that form as a result of fluid-induced melting along a subduction zone (the Wrangell and Garibaldi volcanic belts) , as a result of decompression where the crust is thinning and stretching (Stikine Volcanic Belt and Wells Gray-Clearwater Volcanic Field), and because of mantle plume activity (Anahim Volcanic Belt).

Review Questions

1. What are the three main tectonic settings for volcanism on Earth?
2. What is the primary mechanism for partial melting at a convergent plate boundary?
3. Why are the viscosity and gas content of a magma important in determining the type of volcanic rocks that will be formed when that magma is extruded?
4. Why do the gases in magma not form gas bubbles when the magma is deep within the crust?
5. Where and why do pillow lavas form?
6. What two kinds of volcanic materials make up a composite volcano?
7. What is a lahar, and why are lahars commonly associated with eruptions of composite volcanoes?
8. Under what other circumstances might a lahar form?
9. Why do shield volcanoes have gentle slopes?
10. Which type of volcanic mountain would last longest: a shield volcano, a cinder cone, or a composite volcano?
11. Why is weak seismic activity (small earthquakes) typically associated with the early stages of a volcanic eruption?
12. How can GPS technology be used to help monitor a volcano for activity?
13. What is the likely geological origin of the Nazko Cone?
14. What might be the explanation for southwestern B.C. having much less subduction-related volcanism than adjacent Washington and Oregon?

Answers at the end of the chapter

References

Bressan, D (2012). *Geology Scene Investigation: Death by Volcanic Fire*. Retrieved 27 August 2017 from https://blogs.scientificamerican.com/history-of-geology/geology-scene-investigation-death-by-volcanic-fire

British Geological Survey (n.d.). *Eyjafjallajökull eruption, Iceland | April/May 2010.* Retrieved 27 August 2017 from http://www.bgs.ac.uk/research/volcanoes/icelandic_ash.html

Digital History Project (2011). *"Eyewitness Account to Eruption of Mont Pelee Matinique St Pierre Fort de France" By Ellery S. Scott.* Retrieved 27 August 2017 from http://www.digitalhistoryproject.com/2011/09/eyewitness-account-to-eruption-of-mont.html

Geological Survey of Canada (n.d.) *Catalog of Canadian Volcanoes: Anahim volcanic belt.* Retrieved on 22 September 2017 from https://web.archive.org/web/20080513070258/http://gsc.nrcan.gc.ca/volcanoes/cat/feature_garibaldi_e.php

Geological Survey of Canada (n.d.) *Catalog of Canadian Volcanoes: Garibaldi volcanic belt: Garibaldi Lake volcanic field.* Retrieved on 22 September 2017 from https://web.archive.org/web/20080513070258/http://gsc.nrcan.gc.ca/volcanoes/cat/feature_garibaldi_e.php

Kiser, E., Palomeras, I., Levander, A., Zelt, C., Harder, S., Schmandt, B., Hansen, S., Creager, K., & Ulberg, C. (2016). Magma reservoirs from the upper crust to the Moho inferred from high-resolution Vp and Vs models beneath Mount St. Helens, Washington State, USA. *Geology (44)*6, 411-414. DOI: 10.1130/G37591.1

Lin, G, Amelung, F, Lavallee, Y, and Okubo, P. (2014). Seismic evidence for a crustal magma reservoir beneath the upper east rift zone of Kilauea volcano, Hawaii. *Geology, 42*(3), 187-190. DOI: 10.1130/G35001.1

Regelous, M., Hofmann, A. W., Abouchami, W., & Galer, S. J. G. (2003) Geochemistry of lavas from the Emperor Seamounts, and the geochemical evolution of Hawaiian magmatism from 85 to 42 Ma. *Journal of Petrology 44*(1), 113-140. DOI: 10.1093/petrology/44.1.113

Rosen, J. (2015). *Benchmarks: May 8, 1902: The deadly eruption of Mount Pelée.* Retrieved 27 August 2017 from https://www.earthmagazine.org/article/benchmarks-may-8-1902-deadly-eruption-mount-pelee

Rubin, K. (n.d.). *Mauna Loa Volcano.* Retrieved 23 August 2017 from https://www.soest.hawaii.edu/GG/HCV/maunaloa.html

Skulski, T., Francis, D., & Ludden, J. (1991) Arc-transform magmatism in the Wrangell volcanic belt. *Geology (19)*1, 11-14. doi:10.1130/0091-7613(1991)019<0011:ATMITW>2.3.CO;2

Trop, J. M., Hart, W. K., Snyder, D., & Idleman, B. (2012). Miocene basin development and volcanism along a strike-slip to flat-slab subduction transition: Stratigraphy, geochemistry, and geochronology of the central Wrangell volcanic belt, Yakutat-North America collision zone. *Geosphere (8)*4, 805-834. doi:10.1130/GES00762.1

U. S. Geological Survey (2013) *Mt. St. Helens National Volcanic Monument.* Retrieved on 11 June 2017 from https://volcanoes.usgs.gov/volcanoes/st_helens/st_helens_geo_hist_106.html

U. S. Geological Survey (1997). *Pyroclastic flows.* Retrieved 27 August 2017 from https://pubs.usgs.gov/gip/msh//pyroclastic.html

U. S. Geological Survey, Volcano Hazards Program (n.d.). *Indian Heaven Volcanic Field.* Retrieved 19 September 2017 from https://volcanoes.usgs.gov/volcanoes/indian_heaven/

U. S. Geological Survey, Volcano Hazards Program (n.d.). *Mount St. Helens: 1980 Cataclysmic Eruption.* Retrieved 19 September 2017 from https://volcanoes.usgs.gov/volcanoes/st_helens/st_helens_geo_hist_99.html

Volcanoes Canada, Canadian Hazards Information System, Natural Resources Canada (n.d.). *Where Are Canada's Volcanoes?* Retrieved on 22 September 2017 from http://chis.nrcan.gc.ca/volcano-volcan/can-vol-en.php

Additional Resources

Canadian National Seismograph Network station map:
http://www.earthquakescanada.nrcan.gc.ca/stndon/CNSN-RNSC/stnbook-cahierstn/index-eng.php?tpl_sorting=map&CHIS_SZ=west

Answers to Review Questions

1. The three main tectonic settings for volcanism are (1) subduction zones at convergent plate boundaries, (2) divergent plate boundaries, and (3) mantle plumes (a.k.a. hot spots).
2. The primary mechanism for partial melting at a convergent plate boundary is the addition of water to hot mantle rock. The water reduces the melting temperature of the rock (flux melting).
3. The explosiveness of a volcanic eruption depends on the pressure of the magma. Gases create that pressure, and if the magma is viscous those gases cannot escape easily. Felsic and intermediate magmas tend to have more gas than mafic magmas, and are also more viscous, trapping the gas in.
4. When magma is deep within the crust the pressure is too high for the gases to bubble out of solution.
5. Pillow lavas form where mafic lava erupts in water. When the magma oozes out into the water the outside cools first forming a hard skin that maintains the pillow shape.
6. Composite volcanoes are formed of layers of lava flows and tephra (volcanic fragments ranging from fine ash to blocks and bombs) from explosive eruptions.
7. A lahar is a mud flow or debris flow on a volcano. Lahars are common on composite volcanoes because they are steeper than shield volcanoes, they typically have ice and snow, and they are not as strong as shield volcanoes.
8. Some lahars form during an eruption when snow and ice melt quickly, while others may form from heavy rain.
9. The lava that forms shield volcanoes is typically low viscosity. It can flow easily and also tends to form lava tubes. As a result, it is able to travel a long way from the vent, forming a low broad shield.
10. Cinder cones erode rapidly because they are mostly piles of tephra. Composite volcanoes are more resistant to weathering and erosion than cinder cones because lava flows help hold together tephra, but composite volcanoes still don't last as long as shield volcanoes. Shield volcanoes are stronger because they consist more of lava flows than tephra.
11. Weak seismic activity is associated with all stages of a volcanic eruption. In the early stages magma is moving at depth and pushing rock aside, creating small earthquakes. The flow of magma can also produce special type of seismic response known as a harmonic tremor.
12. GPS technology is used to determine if there is any slow deformation of the flanks of a volcano related to movement of magma toward the surface.
13. The Nazko Cone is thought to be related to a mantle plume.
14. One hypothesis to explain the lower rate of volcanism in British Columbia than in adjacent Washington and Oregon is that the northern part of the Juan de Fuca Plateis not subducting as quickly as the rest of the plate.

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-NC-SA 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 12. Earthquakes

Adapted by Karla Panchuk from *Physical Geology* by Steven Earle

Figure 12.1 Demolition of a structure damaged when an earthquake of magnitude 6.3 struck Christchurch, New Zealand on February 22, 2011. Many structures had already sustained damage from an earthquake that struck six months earlier, in September of 2010. Collapsing structures and falling debris accounted for most of the 185 deaths. *Source: Karla Panchuk (2017) CC BY-SA 4.0. Photograph: Terry Philpott (2012) CC BY-NC 2.0. See Appendix C for more attributions.*

Learning Objectives

After reading this chapter and answering the *Review Questions* at the end, you should be able to:

- Explain how elastic deformation of Earth's crust results in earthquakes.
- Describe how the main shock and immediate aftershocks define the rupture surface of an earthquake, and explain how the transfer of stress to other parts of a fault is related to aftershocks.
- Explain the process of episodic tremor and slip.
- Explain how plate tectonic setting affects where earthquakes occur.
- Distinguish between an earthquake's magnitude and its intensity, and explain how these are determined.
- Describe the hazards caused by earthquakes, including ground shaking, fires, slope failures, liquefaction, and tsunami.
- Explain how the risk of an earthquake can be assessed, and describe steps that governments and individuals can take to minimize the impacts of earthquakes.

Why Study Earthquakes?

On the morning of June 23, 1946, a magnitude 7.3 earthquake struck Vancouver Island. It caused substantial damage to structures, including the school shown in Figure 12.2, and resulting in one fatality. The shaking was so violent that the seismograph measuring the earthquake in Victoria was broken a few seconds after the earthquake started. Even if the seismograph had survived, it was not sensitive enough to provide sufficiently detailed information about local earthquakes, and was unable to record some of them at all. There was only one monitoring station in British Columbia, so while it was possible to determine how far away the earthquake was from the station, geologists were not able to say in which direction. In 1955, Canadian seismologist W. G. Milne wrote that "the 1946 earthquake indicated in a very forceful manner the need for better instruments for the study of earthquakes in British Columbia." A network of improved instruments was established as a direct result of the earthquake.

Figure 12.2 Damage to an elementary school in Courtenay, British Columbia after a magnitude 7.3 earthquake on Sunday, June 23, 1946. Left: A hole left after the chimney collapsed through the roof. Right: Damage inside the school. In addition to damaging structures, the earthquake triggered numerous slope failures. *Source: Photographs courtesy of Earthquakes Canada. See Appendix C for image sources and terms of use.*

Time and time again earthquakes have caused massive damage and many, many casualties. Recording earthquakes and determining their location of origin is important for establishing what geological conditions are responsible for the earthquakes, and understanding the risk they pose. After new, more sensitive instruments were installed at the Dominion Astrophysical Observatory in Victoria, geologists quickly learned that they had underestimated earthquake activity in the area. Between June of 1948 and August of 1951, 224 local earthquakes were recorded!

By studying earthquakes, geoscientists and engineers are making progress toward learning how to minimize earthquake damage, and how to reduce the number of people affected by earthquakes. This knowledge can be communicated to governments, so they are aware of what is needed to keep the population safe. It can also be communicated to individuals, so they know what to expect and do in the event of an earthquake, and can be adequately prepared with emergency supplies.

12.1 What Is An Earthquake?

Earthquake Shaking Comes from Elastic Deformation

Earthquakes occur when rock **ruptures** (breaks), causing rocks on one side of a fault to move relative to the rocks on the other side. Although motion along a fault is part of what happens when an earthquake occurs, rocks grinding past each other is not what creates the shaking. In fact, it could be said that the earthquake happens *after* rocks have undergone most of the displacement. Consider this: if rocks slide a few centimetres or even metres along a fault, would that motion *alone* explain the incredible damage caused by some earthquakes? If you were in a car that suddenly accelerated then stopped, you would feel a jolt. But earthquakes are not a single jolt. Buildings can swing back and forth until they shake themselves to pieces, train tracks can buckle and twist into s-shapes, and roads can roll up and down like waves on the ocean. During an earthquake, rock is not only slipping. It is also vibrating like a plucked guitar string.

Rocks might seem rigid, but when stress is applied they may stretch. If there hasn't been too much stretching, a rock will snap back to its original shape once the stress is removed. Deformation that is

reversible is called **elastic deformation**. Rocks that are stressed beyond their ability to stretch can rupture, allowing the rest of the rock to snap back to its original shape. The snapping back of the rock returning to its original shape causes the rock to vibrate, and this is what causes the shaking during an earthquake. The snapping back is called **elastic rebound**.

Figure 12.3 (top) shows this sequence of events. Stress is applied to a rock and deforms it. The deformed rock ruptures, forming a fault. After rupturing, the rock above and below the fault snaps back to the shape it had before deformation.

Figure 12.3 Elastic deformation, rupture, and elastic rebound. Top: Stress applied to a rock causes it to deform by stretching. When the stress becomes too much for the rock, it ruptures, forming a fault. The rock snaps back to its original shape in a process called elastic rebound. Bottom: On an existing fault, asperities keep rocks on either side of the fault from sliding. Stress deforms the rock until the asperities break, releasing the stress, and causing the rocks to spring back to their original shape. *Source: Karla Panchuk (2017) CC BY 4.0. Modified after Steven Earle (2015) CC BY 4.0*

Ruptures can also occur along pre-existing faults (Figure 12.3, bottom). The rocks on either side of the fault are locked together because bumps along the fault, called asperities, prevent the rocks from moving relative to each other. When the stress is great enough to break the asperities, the rocks on either side of the fault can slide again. While the rocks are locked together, stress can cause elastic deformation. When asperities break and release the stress, the rocks undergo elastic rebound and return to their original shape.

Rupture Surfaces Are Where the Action Happens

Images like 12.3 are useful for illustrating elastic deformation and rupture, but they can be misleading. The rupture that happens doesn't occur as in 12.3, with the block being ruptured through and through. The rupture and displacement only happen along a subsection of a fault, called the **rupture surface**. In Figure 12.4, the rupture surface is the dark pink patch. It takes up only a part of the **fault plane** (lighter pink). The fault plane represents the surface where the fault exists, and where ruptures have happened in the past. Although the fault plane is drawn as being flat in Figure 12.4, faults are not actually perfectly flat.

The location on the fault plane where the rupture happens is called the **hypocentre** or **focus** of the earthquake (Figure 12.4, right). The location on Earth's surface immediately above the hypocentre is the **epicentre** of the earthquake.

Figure 12.4 Rupture surface (dark pink), on a fault plane (light pink). The diagram represents a part of the crust that may be tens or hundreds of kilometres long. The rupture surface is the part of the fault plane along which displacement occurred. Left: In this example, the near side of the fault is moving to the left, and the lengths of the arrows within the rupture surface represent relative amounts of displacement. Coloured arrows represent propagation of failure on a rupture surface. In this case, the failure starts at the dark blue heavy arrow and propagates outward, reaching the left side first (green arrows) and the right side last (yellow arrows). Right: An earthquake's location can be described in terms of its hypocentre (or focus), the location on the fault plane where the rupture happens, or in terms of its epicentre (red star), the location above the hypocentre. *Source: Left- Steven Earle (2015) CC BY 4.0; Right- Karla Panchuk (2017) CC BY 4.0.*

Within the rupture surface, the amount of displacement varies. In Figure 12.4, the larger arrows indicate where there has been more displacement, and the smaller arrows where there has been less. Beyond the edge of the rupture surface there is no displacement at all. Notice that this particular rupture surface doesn't even extend to the land surface of the diagram.

The size of a rupture surface and the amount of displacement along it will depend on a number of factors, including the type and strength of the rock, and the degree to which the rock was stressed beforehand. The magnitude of an earthquake will depend on the size of the rupture surface and the amount of displacement.

A rupture doesn't occur all at once along a rupture surface. It starts at a single point and spreads rapidly from there. Figure 12.4 illustrates a case where rupturing starts at the heavy blue arrow in the middle, then continues through the lighter blue arrows. The rupture spreads to the left side (green arrows), then the right (yellow arrows).

Depending on the extent of the rupture surface, the propagation of **failures** (incremental ruptures contributing to making the final rupture surface) from the point of initiation is typically completed within seconds to several tens of seconds. The initiation point isn't necessarily in the centre of the rupture surface; it may be close to one end, near the top, or near the bottom.

Shifting Stress Causes Foreshocks and Aftershocks

Earthquakes don't usually occur in isolation. There is often a sequence in which smaller earthquakes occur prior to a larger one, and then progressively smaller earthquakes occur after. The largest earthquake in the series is the **mainshock**. The smaller ones that come before are **foreshocks**, and the smaller ones that come after are **aftershocks**. These descriptions are relative, so it can be necessary to reclassify an earthquake. For example, the strongest earthquake in a series is classified as the mainshock, but if another even bigger one comes after it, the bigger one is called the mainshock, and the earlier one is reclassified as an aftershock.

A rupture surface does not fail all at once. A rupture in one place leads to another, which leads to another. Aftershocks and foreshocks represent the same thing, except on a much larger scale. The rupture illustrated in Figure 12.4 reduced stress in one area, but in doing so, transferred stress to others (Figure 12.5). Imagine a frayed rope breaking strand by strand. When a strand breaks, the tension on that strand is released, but the remaining strands must still hold up the same amount of weight. If another strand breaks under the increased burden, the remaining strands have an even greater burden than before. In the same way that the stress causes one strand after another to fail, a rupture can trigger subsequent ruptures nearby.

Figure 12.5 Stress changes related to an earthquake. Stress decreases in the area of the rupture surface, but increases on adjacent parts of the fault. *Source: Steven Earle (2015) CC BY 4.0*

Numerous aftershocks were associated with the magnitude 7.8 earthquake that struck Haida Gwaii in October of 2012 (Figure 12.6; mainshock in red, aftershocks in white). Some of the stress released by the mainshock was transferred to other nearby parts of the fault, and contributed to a cascade of smaller ruptures. But stress transfer need not be restricted to the fault along which an earthquake happened. It will affect the rocks in general around the site of the earthquake and may lead to increased stress on other faults in the region. The aftershocks from the Haida Gwaii earthquake are scattered rather than located only on the main faults.

The effects of stress transfer may not show immediately. Aftershocks can be delayed for hours, days, weeks, or even years. Because stress transfer affects a region, not just a single fault, and because there can be delays between the event that transferred stress and the one that was triggered by the transfer, it can sometimes be hard to be know whether one earthquake is actually associated with another, and whether a foreshock or aftershock should be assigned to a particular mainshock.

Episodic Tremor and Slip

Episodic tremor and slip (ETS) is periodic slow sliding along part of a subduction boundary. It does not produce recognizable earthquakes, but does produce seismic tremor (observed as rapid seismic vibrations on instruments). It was first discovered on the Vancouver Island part of the Cascadia subduction zone by Geological Survey of Canada geologists Herb Dragert and Gary Rogers.

Figure 12.6 Magnitude 7.8 Haida Gwaii earthquake and aftershocks. Mainshock (red circle marks the epicentre) occurred on October 28th, 2012. Aftershocks are for the period from October 28th to November 10th of 2012. Although the epicentre is near a transform boundary, the rupture was influenced more by compression related to the subduction zone. *Source: Karla Panchuk (2017) CC BY 4.0. Base map with epicentres from the U. S. Geological Survey Latest Earthquakes tool. Subduction zone after Wang et al. (2015). See Appendix C for more attributions.*

The boundary between the subducting Juan de Fuca plate and the North America plate can be divided into three segments (Figure 12.7). The cold upper part of the boundary is the **locked zone**. There the plates are stuck together for long periods of time. When slip does occur, it generates very large earthquakes. The last time the locked zone along Canada's west coast slipped was January 26, 1700. It caused an earthquake of magnitude 9. The warm lower part of the boundary, called the **continuous slip zone**, is sliding continuously because the warm rock is weaker. The central part of the boundary, the **ETS zone**, isn't cold enough to be stuck, but isn't warm enough to slide continuously. Instead it slips episodically approximately every 14 months for about 2 weeks, moving a few centimetres each time.

Figure 12.7 Episodic tremor and slip along the Cascadia subduction zone. The Juan de Fuca plate is locked to the North American plate at the top of the subduction zone, but lower down it is slipping continuously. In the intermediate (ETS) zone, the plate alternately sticks and slips on a regular schedule. *Source: Steven Earle (2015) CC BY 4.0*

It might seem that periodic slip along this part of the plate helps to reduce tension, and thus reduce the risk of a large earthquake. In fact, the opposite is likely the case. The movement along the ETS part of the plate boundary transfers stress to the adjacent locked part of the plate. During the two-week ETS period, the transfer of stress means an increased chance of a large earthquake.

Since 2003, ETS processes have also been observed in subduction zones in Mexico and Japan.

12.2 Measuring Earthquakes

The shaking from an earthquake travels away from the rupture in the form of **seismic waves**. Seismic waves are measured to determine the location of the earthquake, and to estimate the amount of energy released by the earthquake (its **magnitude**).

Types of Seismic Waves

Seismic waves are classified according to where they travel, and how they move particles.

Body Waves

Seismic waves that travel through Earth's interior are called **body waves**. **P-waves** are body waves that move by alternately compressing and stretching materials in the direction the wave moves. For this reason, P-waves are also called compression waves. The "P" in P-wave stands for primary, because P-waves are the fastest of the seismic waves. They are the first to be detected when an earthquake happens.

A P-wave can be simulated by fixing one end of a spring to a solid surface, then giving the other end a sharp push toward the surface (Figure 12.8, top). The compression will propagate (travel) along the length of the spring. Some parts of the spring will be stretched, and others compressed. Any one point on the spring will jiggle forward and backward as the compression travels along the spring.

Figure 12.8 Seismic waves simulated using a spring and rope attached to a fixed surface. Top: P-waves travel as pulses of compression. Bottom: S-waves move particles at right angles to the direction of motion. *Source: Karla Panchuk (2018) CC BY 4.0 modified after Steven Earle (2015) CC BY 4.0*

S-waves are body waves that move with a shearing motion, shaking particles from side to side. S-waves can be simulated by fixing one end of a rope to a solid surface, then giving the other end a flick (Figure 12.8, bottom). Any one point on the rope will move from side to side at a right angle to the direction in which the snaking motion is traveling. The "S" in S-wave stands for secondary, because S-waves are slower than P-waves, and are detected after the P-waves are measured. S-waves cannot travel through liquids.

P-waves and S-waves can travel rapidly through geological materials, at speeds many times the speed of sound in air.

Surface Waves

When body waves reach Earth's surface, some of their energy is transformed into surface waves, which travel along Earth's surface. Two types of surface waves are **Rayleigh waves** and **Love waves** (Figure

12.9). Rayleigh waves (R-waves) are characterized by vertical motion of the ground surface, like waves rolling on water. Love waves (L-waves) are characterized by side-to-side motion. Notice that the effects of both kinds of surface waves diminish with depth in Figure 12.9.

Surface waves are slower than body waves, and are detected after the body waves. Surface waves typically cause more ground motion than body waves, and therefore do more damage than body waves.

Figure 12.9 Surface waves travel along Earth's surface and have a diminished impact with depth. Rayleigh waves (left) cause a rolling motion, and Love waves (right) cause the ground to shift from side to side. *Source: Steven Earle (2015) CC BY 4.0. See Appendix C for more attributions.*

Recording Seismic Waves Using a Seismograph

A **seismometer** is an instrument that detects seismic waves. An instrument that combines a seismometer with a device for recording the waves is called a **seismograph**. The graphical output from a seismograph is called a **seismogram**.

A seismograph consists of a frame or housing that is firmly anchored to the ground (Figure 12.10). A mass is suspended from the housing, and can move freely on a spring. When the ground shakes, the housing shakes with it, but the mass remains fixed. A pen attached to the mass moves up and down on a rotating drum of paper, drawing a wavy line, the seismogram. The seismograph in Figure 12.10 (right) is oriented to measure vertical ground motion. The photo on the left shows a seismograph oriented to record horizontal ground motion.

Figure 12.10 How a seismograph records earthquakes. *Source: Left- Karla Panchuk (2018) CC BY-NC-SA 4.0 modified after IRIS (2012) "How Does a Seismometer Work?"; Right- Karla Panchuk (2018) CC BY-SA 4.0, photo by Z22 (2014) CC BY-SA 3.0. See Appendix C for more attributions.*

The pen and drum of a mechanical seismograph record the motion of the ground relative to the mass. However, unless an earthquake causes a large amount of ground motion directly beneath the seismograph, the height of the wave recorded on paper might be very small, making the seismogram difficult to analyze. The seismograph on the right has a device to amplify the ground motion, drawing larger waves that are easier to study.

Modern seismographs record shaking as electrical signals, and are able to transmit those signals. This means seismologists need not return to the instrument to collect recordings before the records can be examined.

Finding The Location of an Earthquake

P-waves travel faster than S-waves. As the waves travel away from the location of an earthquake, the P-wave gets farther and father ahead of the S-wave. Therefore, the farther a seismograph is from the location of an earthquake, the longer the delay between when the P-wave arrival is recorded, and the S-wave arrival is recorded. The delay between the P-wave and S-wave arrival appears as a widening gap in a diagram of P-wave and S-wave travel times (Figure 12.11, grey lines).

P-wave and S-wave arrival times can be identified on seismograms. In the three seismograms in Figure 12.11, the arrivals of the P-waves and S-waves are marked with arrows, and the interval in minutes between the P-wave and S-wave arrivals are noted. The seismograms were recorded at three different **seismic stations** (earthquake monitoring locations equipped with seismographs). The distance of each station from the earthquake is determined by finding the distance along the graph where the gap between the P-wave and S-wave travel-time curves matches the delay between P-wave and S-wave arrivals on the seismogram.

Figure 12.11 Using P-wave and S-wave travel times to determine how far seismic waves have travelled. Grey curves show the distance travelled by P-waves and S-waves after an earthquake occurs. P-waves are faster than S-waves, and the gap between them increases with time and distance. The delays between P-wave and S-wave arrivals on seismograms are matched to the curve to find the distances of seismic stations from the source of the seismic waves. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0 modified after IRIS (n.d.) "How Are Earthquakes Located?"*

The delay between the P-wave and S-wave arrival at a seismic station can indicate how far the station is from the source of the earthquake, but not the direction from which the seismic waves travelled. The possible locations of the earthquake can be represented on a map by drawing a circle around the seismic station, with the radius of the circle being the distance determined from the P-wave and S-wave travel times (Figure 12.12). If this is done for at least three seismic stations, the circles will intersect at the origin of the earthquake.

Figure 12.12 Locating earthquakes by drawing three circles with radii of lengths determined from P-wave and S-wave travel times. Station names (SOCO, TEIG, SSPA) correspond to seismograms in Figure 12.11. *Source: IRIS (n.d.) "How Are Earthquakes Located?" See Appendix C for terms of use.*

How Big Was It?

Earthquakes can be described in terms of their **magnitude**, which reflects the amount of energy released by the shaking. They can also be described in terms of **intensity**, which characterizes the impact of the shaking on people and their surroundings.

Earthquake Magnitude

Earthquake magnitudes are determined by measuring the amplitudes of seismic waves. The **amplitude** is the height of the wave relative to the baseline (Figure 12.13). Wave amplitude depends on the amount of energy carried by the wave. The amplitudes of seismic waves reflect the amount of energy released by earthquakes.

Figure 12.13 Seismogram for a small earthquake that occurred near Vancouver Island in 1997. The maximum amplitude of the S-wave is indicated. *Source: Karla Panchuk (2018) CC BY 4.0 modified after Steven Earle (2015) CC BY 4.0*

The **Richter magnitude** scale uses the amplitudes of S-waves, and corrects for the decrease in amplitude that happens as the waves travel away from their source. The correction depends on how seismic waves interact with the specific rock types through which they travel, and therefore on local conditions, so the Richter magnitude is also referred to as the **local magnitude**.

While news reports about earthquakes might still refer to the "Richter scale" when describing magnitudes, the number they report is most likely the **moment magnitude**. The moment magnitude is calculated from the **seismic moment** of an earthquake. The seismic moment takes into account the surface area of the region that ruptured, how much displacement occurred, and the stiffness of the rocks. Moment magnitude can capture the difference between short earthquakes and longer ones resulting from larger ruptures, even

of both types of earthquakes generate the same amplitude of waves. The moment magnitude scale is also better for earthquakes that are far from the seismic station. Seismic wave measurements are still used to determine the moment magnitude, however different waves are used than for the local magnitude scale.

The magnitude scale is a logarithmic one rather than a linear one- an increase of one unit of magnitude corresponds to a 32 times increase in energy release (Figure 12.14). There are far more low-magnitude earthquakes than high-magnitude earthquakes. In 2017 there were 7 earthquakes of M7 (magnitude 7) or greater, but millions of tiny earthquakes.

Figure 12.14 Earthquake magnitude and corresponding energy release. Energy release increases by approximately 32 times for each unit change in magnitude. *Source: IRIS (n.d.) "How Often Do Earthquakes Occur?"*

Earthquake Intensity

Intensity scales were first used in the late 19th century, and then adapted in the early 20th century by Giuseppe Mercalli and modified later by others to form what we now call the **Modified Mercalli Intensity Scale** (Table 12.1). To determine the intensity of an earthquake, reports are collected about what people felt and how much damage was done. The reports are then used to assign intensity ratings to regions where the earthquake was felt.

Table 12.1 Modified Mercalli Intensity Scale

I	Not felt	Not felt except by a very few under especially favourable conditions
II	Weak	Felt only by a few persons at rest, especially on upper floors of buildings
III	Weak	Felt quite noticeably by persons indoors, especially on upper floors of buildings; many people do not recognize it as an earthquake; standing motor cars may rock slightly; vibrations similar to the passing of a truck; duration estimated
IV	Light	Felt indoors by many, outdoors by few during the day; at night, some awakened; dishes, windows, doors disturbed; walls make cracking sound; sensation like heavy truck striking building; standing motor cars rocked noticeably
V	Moderate	Felt by nearly everyone; many awakened; some dishes, windows broken; unstable objects overturned; pendulum clocks may stop
VI	Strong	Felt by all, many frightened; some heavy furniture moved; a few instances of fallen plaster; damage slight
VII	Very Strong	Damage negligible in buildings of good design and construction; slight to moderate in well-built ordinary structures; considerable damage in poorly built or badly designed structures; some chimneys broken
VIII	Severe	Damage slight in specially designed structures; considerable damage in ordinary substantial buildings with partial collapse; damage great in poorly built structures; fall of chimneys, factory stacks, columns, monuments, walls; heavy furniture overturned
IX	Violent	Damage considerable in specially designed structures; well-designed frame structures thrown out of plumb; damage great in substantial buildings, with partial collapse; buildings shifted off foundations
X	Extreme	Some well-built wooden structures destroyed; most masonry and frame structures destroyed with foundations; rails bent
XI	Extreme	Few, if any (masonry), structures remain standing; bridges destroyed; broad fissures in ground; underground pipelines completely out of service; earth slumps and land slips in soft ground; rails bent greatly
XII	Extreme	Damage total; waves seen on ground surfaces; lines of sight and level distorted; objects thrown upward into the air

Source: U. S. Geological Survey (1989). *The Severity of an Earthquake*. USGS General Interest Publication 1989-288-913 https://pubs.usgs.gov/gip/earthq4/severitygip.html

Intensity values are assigned to locations, rather than to the earthquake itself. This means that intensity can vary for a given earthquake, depending on the proximity to the epicentre and local conditions. For the 1946 M7.3 Vancouver Island earthquake, intensity was greatest in the central island region (Figure 12.15). In some communities within this region, chimneys were damaged on more than 75% of buildings. Some roads were made impassable, and a major rock slide occurred. The earthquake was felt as far north as Prince Rupert, as far south as Portland Oregon, and as far east as the Rockies, but with less intensity.

Intensity estimates are important as a way to identify regions that are especially prone to strong shaking. A key factor is the nature of the underlying geological materials. The weaker the underlying materials, the more likely it is that there will be strong shaking. Areas underlain by strong solid bedrock tend to experience far less shaking than those underlain by unconsolidated river or lake sediments.

An example of this effect is the 1985 M8 earthquake that struck the Michoacán region of western Mexico, southwest of Mexico City. There was relatively little damage near the epicentre, but 350 km away in heavily populated Mexico City there was tremendous damage and approximately 5,000 deaths. The reason is that Mexico City was built largely on the unconsolidated and water-saturated sediment of former Lake Texcoco. These sediments resonate at a frequency of about two seconds, which was similar to the frequency of the body waves that reached the city. Consequently, the shaking was amplified. Survivors of the disaster recounted that the ground in some areas moved up and down by approximately 20 cm every two seconds for over two minutes. Damage was greatest to buildings between 5 and 15 storeys tall, because they also resonated at around two seconds, which amplified the shaking.

Figure 12.15 Intensity map for the M7.3 Vancouver Island earthquake on June 23, 1946. *Source: Earthquakes Canada, Natural Resources Canada (2016). See Appendix C for terms of use.*

12.3 Earthquakes and Plate Tectonics

Bands of earthquakes trace out plate boundaries (coloured dots, Figure 12.16). The depths of earthquakes, and the width of the band, depend on the type of plate boundary. Mid-ocean ridges and transform margins have shallow earthquakes (usually less than 30 km deep), in narrow bands close to plate margins. Subduction zones have earthquakes at a range of depths, including some more than 700 km deep. Bands of earthquakes are wider along subduction zones because they take place throughout the subducting slab that extends beneath the opposing plate. Wide swaths of scattered earthquakes may correspond to continent-continent collision zones, such as between the Eurasian plate and the African, Arabian, and Indian plates to the south. Wide swaths of scattered earthquakes may also correspond to continental rift zones, such as in eastern Africa.

Earthquakes are also relatively common at a few locations away from plate boundaries. Some are related to the buildup of stress due to continental rifting or the transfer of stress from other regions, and some are not well understood. Locations include the Great Rift Valley area of Africa, the Lake Baikal area of Russia, and Tibet.

Earthquakes at Divergent and Transform Plate Boundaries

Earthquakes along divergent and transform plate margins are shallow (usually less than 30 km deep) because below those depths, rock is too hot and weak to avoid being permanently deformed by the stresses in those settings. If deformation is permanent, then removing the stress does not result in the rocks snapping back to their original shape. No snapping back means no shaking.

Figure 12.16 Earthquakes greater than magnitude 5, from 2000 to 2008. Bands of earthquakes mark tectonic plates. Narrow bands with shallow earthquakes (marked in red) indicate transform boundaries or mid-ocean ridge divergent boundaries. Wider bands with earthquakes at a range of depths are subduction zones. Wide bands of scattered earthquakes mark continent-continent convergent margins (e.g., between the Indian and Eurasian plates), or continental rift zones (e.g., in eastern Africa). *Source: Lisa Christiansen, Caltech Tectonics Observatory (2008). (Plate and ocean basin labels added.) See Appendix C for terms of use.*

Mid-ocean ridge divergent plate margins are offset by numerous transform faults (Figure 12.17). The locations of earthquakes along mid-ocean ridges, and the mechanisms for causing them, depend on how rapidly the mid-ocean ridges are spreading.

Figure 12.17 Locations of earthquakes of magnitude 4 and greater from 1990 to 2010 along two mid-ocean ridges. Plate boundaries are marked in red. Arrows show the direction of plate motion. Left: Rapidly spreading Pacific-Antarctic ridge with earthquakes concentrated along transform faults. Right: Slowly spreading Southwest Indian Ridge, with earthquakes along both spreading segments and transform faults. *Source: Karla Panchuk (2017) CC BY 4.0. Base maps with epicentres generated using the U. S. Geological Survey Latest Earthquakes website.*

The Pacific-Antarctic Ridge (left) is spreading relatively rapidly at 42 to 94 mm/year, depending on the location along the ridge. Rapid spreading causes rocks near the axis of the spreading centre to be hot and weak. As a result, most of the earthquakes (white dots) are located along transform faults, where rocks are cooler and stronger. Along rapidly spreading ridges, new ocean crust is bent upward into wide, high ridges. As spreading proceeds and crust moves away from the ridge, the bend is relaxed, and the crust stretches and breaks. This triggers earthquakes many kilometres away from the ridge.

The Southwest Indian Ridge (right) spreads very slowly, at approximately 14 mm/year. Rocks are cooler and stronger along the slowly spreading ridge than along the rapidly spreading one. In the slow-spreading environment, earthquakes are generated when rocks along the ridge axis stretch and break. Earthquakes are more evenly distributed between divergent and transform segments of the boundary than they are along fast-spreading ridges.

Earthquakes in continental rift zones are also shallow, but scattered more broadly than those along mid-ocean ridges. Lake Baikal (Figure 12.28), the world's oldest, deepest, and largest freshwater lake, formed 25 million years ago because of continental rifting. Note the scale in Figure 12.18, and compare how widely the earthquakes (blue dots) are spread in the Lake Baikal region, versus along the mid-ocean ridges in Figure 12.17.

Figure 12.18 Blue circles mark the locations of earthquakes of M4 and greater from 1990 to 2010 along the Lake Baikal rift zone. White lines show some of the faults in the region. White lines with tick marks are normal faults related to spreading. Arrows show the direction of spreading. White lines without tick marks are transform faults. The Siberian Craton (shaded region) is strong 2 billion year old crust. *Source: Karla Panchuk (2017) CC BY 4.0. Faults after U. S. Geological Survey (see references). Base maps (inset and rift views) with epicentres generated using the U. S. Geological Survey Latest Earthquakes website.*

One reason for the difference in earthquake distribution in continental rift zones is that the rifts are only beginning to form. Faulting is "disorganized" within the continental crust. There is no well-established spreading centre, unlike mid-ocean ridges. Another reason is that the locations of faults, and thus earthquakes, in continental rift zones are affected by pre-existing geological structures within continental crust. In the case of the Lake Baikal rift, the strong, ancient crust of the Siberian Craton influences the orientation of the faults forming the rift. Faults run parallel to the craton near Lake Baikal. As rifting extends to the east, the part of the craton in the upper right of Figure 12.18 may deflect rifting southward.

Earthquakes at Convergent Boundaries

Subduction Zones

Along convergent plate margins with subduction zones, earthquakes range from shallow to depths of up to 700 km. Earthquakes occur where the two plates are in contact, as well as in zones of deformation on the

overriding plate, and along the subducting slab deeper within the mantle. The result is that epicentres of earthquakes farther to the interior of the overriding plate will correspond to increasingly deep earthquakes.

Where the Pacific plate subducts beneath the North American plate, forming the Aleutian volcanic arc (Figure 12.19), earthquakes increase in depth moving northward, following the Pacific plate into the mantle. Earthquakes between 0 and 33 km deep (red circles) occur closest to the subduction zone (red line; teeth point in the direction of the subducting slab). While there is some overlap, earthquakes between 33 and 70 km deep (white circles) occur in a band that reaches farther the north. Farthest north are the epicentres for earthquakes between 70 and 300 km deep (green dots). The deepest earthquake during the seven year interval shown in Figure 12.19 is represented by the large green dot farthest to the north. It occurred at a depth of 265 km.

Figure 12.19 Earthquakes of M4.5 and greater from 2010 to 2017 along the Aleutian Trench subduction zone (red line; teeth point in the direction of the subducting slab). White arrows show the directions of plate movement. Circle colours indicate the depths of earthquakes (see legend, lower left). Earthquakes become deeper moving north from the subduction zone. *Source: Karla Panchuk (2017) CC BY 4.0. Base maps with epicentres generated using the U. S. Geological Survey Latest Earthquakes website.*

Earthquakes occur in subduction zones for a variety of reasons. Stresses associated with the collision of two plates cause deformation in the overriding plate, and thus shallow earthquakes. Shallow earthquakes also happen on the subducting slab when a locked zone (orange line, Figure 12.20) ruptures. The locked zone is where the largest earthquakes on Earth, called **megathrust earthquakes**, occur. There is the potential for a wider rupture zone on a gently dipping subduction zone boundary compared to other boundaries.

If subduction is rapid, the subducting plate will bend more as it enters the mantle (slab A in Figure 12.20), causing the upper edge of the plate to stretch, and the interior and lower edge to be compressed. Stress from bending can cause shallow to intermediate earthquakes on these plates. Even without bending, the subducting slab can become stretched by its own weight as it falls into the mantle.

The 410 km and 660 km discontinuities in Figure 12.20 mark boundaries where minerals transform into other, denser minerals that are stable at higher pressures and temperatures. When the subducting slab reaches the 660 km discontinuity (the top of the lower mantle), the increase in density in the surrounding mantle may slow down the leading edge of the sinking slab. Earthquakes can be generated when the slab is compressed by the lower mantle resisting its motion at the same time that the upper part of the slab continues to fall.

Figure 12.20 Factors contributing to earthquakes in subduction zones. Not all factors shown here are present in all subduction zones. Stress from bending, flexing, and stretching may cause ruptures. Changes in the mechanical properties of the mantle may affect how subucting slabs move, contributing additional stresses. The histogram at right shows the global average number of earthquakes at depth, for earthquakes greater than M5. The increase in earthquakes beneath 480 km may be caused in part by weakening as olivine transforms into high pressure/temperature polymorphs. *Source: Karla Panchuk (2017) CC BY 4.0. Modified after Green et al. (2010).*

Slower rates of subduction mean that the subducting slab will enter the mantle at a lower angle (slab B in Figure 12.20). These slabs might not have earthquakes from being bent downward into the mantle, as with slab A, but earthquakes may be triggered by changes in stress if the plate relaxes and unbends.

The bar chart on the right of Figure 12.20 shows global average number of earthquakes that occur at different depths. Earthquakes are most abundant at the surface, and then fall to a minimum at 300 km. The number of earthquakes remains low until almost 500 km depth, and reaches a second peak around 600 km depth. The second peak might be explained by interactions between the subducting plate and the dense mantle beneath the 660 km discontinuity, but another hypothesis is that it is related to delayed mineral transformations. The subducting slab warms as it goes deeper into the mantle, but the warming is not uniform. The outer edges of the slab will warm before the interior does. The 410 km discontinuity is where olivine is transformed into the minerals wadsleyite and ringwoodite under normal mantle pressure and temperature conditions. However, if the interior of the subducting slab is still too cool at that depth, olivine will be retained to depths below 410 km. Olivine weakens prior to transforming into the high pressure minerals, and the weakening may make it easier for the slab to rupture.

Continent-Continent Convergence Zones

Where continents collide, earthquakes are scattered over a much wider area compared to earthquakes along mid-ocean ridges, transform margins, or subduction zones. An example is where the Indian plate collides with the Eurasian plate (Figure 12.21). At one time, India was a separate continent, and ocean crust separated India from the Eurasian plate. For a time, a subduction zone existed where ocean lithosphere from the Indian plate subducted beneath the Eurasian plate. But when the two land masses finally met, they became locked together and the subduction zone was closed off. Today the Indian plate is still pushing against the Eurasian plate in the regions indicated by the red arrows in Figure 12.21. The collision is accommodated by transform boundaries along the Indian plate. Regions of overall transform motion are indicated in Figure 12.21 with blue arrows.

Figure 12.21 Earthquakes of M4.5 and greater from 1990 to 2017 along the collision zone between the Indian and Eurasian plates. Red lines- plate boundaries; red arrows- collision zones; blue arrows- transform zones. *Source: Karla Panchuk (2017) CC BY 4.0. Base maps with epicentres generated using the U. S. Geological Survey Latest Earthquakes website.*

The majority of earthquakes in Figure 12.21 occur at depths less than 70 km, however they are still abundant down to 150 km, and extend to more than 300 km depth at some locations. Deeper earthquakes may be caused by continued northwestward subduction of part of the Indian plate beneath the Eurasian plate in this area. Even though the area is no longer a subduction zone, the subducted slab still remains, and is subject to stresses that can trigger earthquakes.

Some of the earthquakes in Figure 12.21 are related to the transform faults on either side of the Indian plate, and most of the others are related to the squeezing caused by the continued convergence of the Indian and Eurasian plates. That squeezing has caused the Eurasian plate to be thrust over the Indian plate, building the Himalayas and the Tibet Plateau to enormous heights. Most of the earthquakes of Figure 12.21 are related to the thrust faults shown in Figure 12.22 (and to hundreds of other similar ones that cannot be shown at this scale). The southernmost thrust fault in Figure 12.22 (the Main Boundary Fault) is equivalent to the convergent boundary in Figure 12.21.

Intraplate Earthquakes

Intraplate earthquakes (within-plate earthquakes) are those that occur away from plate boundaries. Some intraplate earthquakes are related to human activities. When humans trigger earthquakes it is referred to as **induced seismicity**. In Saskatchewan there have been 20 earthquakes since 1985 (all less than magnitude 4), and the majority occurred near potash mines. Excavation changes the stress in surrounding rocks, so earthquakes may occur in the rocks above excavated parts of the mine. In Alberta, induced seismicity is triggered by hydraulic fracturing operations when water pressure increases along existing faults, causing them to rupture.

Figure 12.22 Schematic diagram of the India-Asia convergent boundary, showing examples of the types of faults along which earthquakes are focused. *Source: Steven Earle (2015) CC BY 4.0 after D. Vuichard (Figure 2.3) in Ives and Messerli (1989).*

Intraplate earthquakes not related to human activities often occur along ancient rift zones. In eastern Canada, the Charlevoix Seismic Zone (approximately 100 km northeast of Québec City; Figure 12.23), is associated with a rift-zone faults that developed when an ancient ocean basin began to form more than 500 million years ago.

Figure 12.23 Charlevoix seismic zone, site of intraplate earthquakes. The location of the Charlevoix seismic zone is indicated by the star on the map of Canada. Dots indicate earthquake epicentres. The size of the dots indicates magnitude. White lines indicate fault segments. The dashed circle marks the edge of a crater formed by a meteorite impact 342 million years ago. *Source: Karla Panchuk (2017) CC BY-SA 4.0. Epicentre data from Earthquakes Canada. See Appendix C for more attributions and data sources.*

Coincidentally, the rocks of the Charlevoix seismic zone are also fragmented because of a meteorite impact (the crater margin is indicated by the blue circle in Figure 12.23), weakening them further. While the Charlevoix zone is far from any boundary of the North American plate, tectonic forces acting on plate boundaries are still transmitted to the interior of the continent, contributing to the stress that causes the faults along the rift zone to rupture.

Intraplate earthquakes can be large earthquakes. The Charlevoix seismic zone has had five earthquakes of magnitudes between 6 and 7 since 1663. The New Madrid seismic zone in the Mississippi River Valley had a series of four earthquakes with magnitudes between 7 and 8 in the winter of 1811-1812. The population

of the region was sparse at the time, but today there are major cities in the New Madrid seismic zone, including Memphis, Tennessee, and St. Louis, Missouri.

12.4 The Impacts of Earthquakes

Earthquakes can have direct impacts, such as structural damage to buildings from shaking, and secondary impacts, such as triggering landslides, fires, and tsunami. The types and extent of impacts will depend on local conditions where the earthquake strikes. The geological materials in the area matter, as does the type of terrain, and whether the region is near the coast or not. The extent of impact and type of damage will depend on whether the area is predominantly urban or rural, densely or sparsely populated, highly developed or underdeveloped. It will depend on whether the infrastructure has been designed to withstand shaking.

Damage to Structures from Shaking

As with the example of the 1985 Mexico earthquake, the geological foundations on which structures are built will affect the amount of shaking that occurs. Earthquakes produce seismic waves that vibrate at different rates, or **frequencies**. Waves with rapid vibrations have a high frequency, and waves with slower vibrations have lower frequencies.

The energy of the higher frequency waves tends to be absorbed by solid rock. Lower frequency waves pass through solid rock without being absorbed, but are absorbed and amplified by soft sediments. It is therefore very common to see much worse earthquake damage in areas underlain by soft sediments than in areas of solid rock. During the 1989 Loma Prieta earthquake, parts of a two-layer highway in the Oakland area near San Francisco collapsed where they were built on soft sediments (Figure 12.24).

Figure 12.24 A collapsed section of the Cypress Freeway in Oakland California. *Source: U. S. Geological Survey (1989) Public Domain*

Building damage is also greatest in areas of soft sediments, and multi-storey buildings tend to be more seriously damaged than smaller ones. Buildings can be designed to withstand most earthquakes, and this practice is increasingly applied in earthquake-prone regions. Turkey is one such region, but even though Turkey had a relatively strong building code in the 1990s, adherence to the code was poor. Builders did whatever they could to save costs, including using inappropriate materials in concrete, and reducing the amount of steel reinforcing. The result was more than 17,000 deaths in the 1999 M7.6 Izmit earthquake (Figure 12.25). After two devastating earthquakes that year, Turkish authorities strengthened the building code further, but the new code has been applied only in a few regions, and enforcement of the code is still weak, as revealed by the amount of damage from a M7.1 earthquake in eastern Turkey in 2011.

Figure 12.25 Damage from the 1999 M7.6 Izmit, Turkey earthquake. *Source: Left; USGS (1999) Public Domain; Right: USGS (1999) Public Domain*

Structures underlain by sediments may be at risk of another hazard, called **liquefaction**, in which sediment is transformed into a fluid. When water-saturated sediments are shaken, the grains may lose contact with each other, and no longer support one another. Water between the grains holds them apart, causing the sediment to turn to mud and flow. The loss of support can lead to the collapse of buildings or other structures that might otherwise have sustained little damage. During the 1964 M7.6 earthquake in Niigata, Japan, liquefaction caused buildings to sink into the sediments (Figure 12.26).

Figure 12.26 Collapsed apartment buildings in the Niigata area of Japan. The material beneath the buildings was liquefied by the 1964 earthquake. *Source: DOC/NOAA/NESDIS/NCEI (1964) Public Domain*

Parts of the Fraser River delta are also prone to liquefaction-related damage. The region is characterized by a 2 m to 3 m thick layer of fluvial silt and clay above a layer of water-saturated fluvial sand that is at least 10 m thick (Figure 12.27). Under these conditions, seismic shaking can be amplified, and the sandy sediments will liquefy. This could lead to subsidence and tilting of buildings. Liquefaction can also contribute to slope failures and to fountains of sandy mud (sand volcanoes) in areas where there is loose saturated sand beneath a layer of more cohesive clay. Current building-code regulations in the Fraser delta area require that measures be taken to strengthen the ground beneath multi-story buildings prior to construction.

Figure 12.27 Recent unconsolidated sedimentary layers in the Fraser River delta area (top) and the potential consequences in the event of a damaging earthquake. *Source: Steven Earle (2015) CC BY 4.0*

Experiments with Liquefaction

To see liquefaction for yourself, go to a sandy beach and find a place near the water's edge where the sand is wet. While standing in one place on a wet part of the beach, start moving your feet up and down at a frequency of about once per second. Within a few seconds the previously firm sand will start to lose strength, and you'll gradually sink in up to your ankles.

Figure 12.28 Fine, damp sand before shaking (left) and after (right). *Source: Steven Earle (2015) CC BY 4.0*

If you can't get to a beach, put some sand into a small container, saturate it with water, and then pour the excess water off. Shake the container gently to get the water to separate and then pour the excess water away. You may have to do this more than once. Place a small rock on the surface of the sand. It should sit there for hours without sinking in. Now, holding the container in one hand gently thump the side or the bottom with your other hand, about twice a second. The rock should gradually sink in as the sand around it becomes liquefied.

As you were moving your feet up and down or thumping the container, it's likely that you soon discovered the most effective rate for getting the sand to liquefy. Stepping up and down as fast as you can (several times per second) on the wet beach would not have been effective, nor would you have achieved much by stepping once every several seconds. The body of sand vibrates most readily in response to shaking that is close to its natural harmonic frequency, and liquefaction is also most likely to occur at that frequency.

Figure 12.29 A slope gives way in a suburb of San Salvador after the January 2001 earthquake offshore of El Salvador. This is one of hundreds of slope failures resulting from the earthquake. *Source: U. S. Geological Survey (2001) Public Domain*

Slope Failure

Ground shaking during an earthquake can be enough to weaken rock and loose materials to the point of failure. Earthquakes can also trigger failures on slopes that are already weak. In January of 2011 a M7.6 earthquake offshore of El Salvador triggered slope failures that killed nearly 600 people (Figure 12.29).

Fires

Fires are commonly associated with earthquakes because gas lines rupture and electrical lines are damaged when the ground shakes. Most of the damage in the great 1906 San Francisco earthquake was caused by massive fires in the downtown area of the city (Figure 12.30). Some 25,000 buildings were destroyed by those fires, which were fueled by gas leaking from broken pipes. Fighting the fires was difficult because water mains had also ruptured. Today the risk of fires can be reduced through P-wave early warning systems if utility operators can decrease pipeline pressure and break electrical circuits.

Figure 12.30 Fires in San Francisco following the 1906 earthquake. *Source: Pillsbury Picture Co. (1906) Public Domain, courtesy of the Library of Congress Prints and Photographs Division*

Tsunami

Large earthquakes that take place beneath the ocean have the potential to displace large volumes of water. In a subduction zone, for example, the overriding plate becomes distorted by elastic deformation. It is squeezed laterally and pushed up (Figure 12.31 top). When an earthquake happens, the plate rebounds over an area of thousands of square kilometres, generating waves- a **tsunami** (Figure 12.31, middle). The waves spread across the ocean at velocities of several hundred kilometres per hour. Tsunami can make it to the far side of an ocean in about the same time as a passenger jet.

Tsunami waves gain their height as they travel through shallower waters. In the deep ocean, the waves may be so small as to go undetected by ships, but when they are slowed down by interacting with the ocean floor, they can become much taller (Figure 12.31, bottom). In the tsunami following the 2004 Sumatra earthquake, the tallest waves were more than 30 m high.

Subduction earthquakes must be large to generate significant tsunami. Earthquakes with magnitude less than 7 do not typically generate significant tsunami because the amount of vertical displacement of the sea floor is minimal. Sea-floor transform earthquakes, even large ones (M7 to M8), don't typically generate tsunami either, because the motion is mostly side to side, not vertical.

Figure 12.31 Tsunami triggered along a subduction zone. Top: The overriding crust is deformed because it is locked to the subducting slab. Middle: When the locked zone ruptures, the crust rebounds, and waves are created. Bottom: Tsunami waves have small amplitudes in the deep ocean, but once in shallow water, they slow down, causing the waves to become taller and closer together. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Top and middle modified after Steven Earle (2015) CC BY 4.0. Bottom modified after COMET/UCAR (1997-2017). See Appendix C for COMET/UCAR attribution and terms of use.*

Tsunami can have an impact across an entire ocean basin. They spread across the ocean at velocities of several hundred kilometres per hour, and can make it to the far side of an ocean in about the same time as a passenger jet. In 1700 a rupture along the Cascadia thrust fault running from Vancouver Island to northern California resulted in a M9 earthquake. It generated a tsunami that travelled across the Pacific Ocean, and was recorded in Japan nine hours later. A computer simulation of the tsunami (Figure 12.32) shows how long it took tsunami waves from the Cascadia earthquake to travel across the Pacific Ocean, and how high the waves were. Notice that over all, the waves decrease in height moving away from the rupture, but they increase in height again as they reach the opposite side of the Pacific Ocean.

Figure 12.32 Computer simulation of the tsunami from the 1700 M9 Cascadia earthquake. Colours show open-ocean wave heights. Contours show travel time in hours. Tsunami wave heights increase as the tsunami reached the western margin of the Pacific ocean. *Source: NOAA/PMEL/Center for Tsunami Research (2011) Public Domain*

In spite of the damage that lava flows can cause, they are not the volcanic hazard with the greatest impact on lives and infrastructure. Even the relatively free-flowing Hawai'ian basaltic lava moves slowly enough that it can be escaped on foot. Far more dangerous hazards are related to gases and volcanic debris. However, the largest impact and the greatest suffering are caused not by the immediate effects of volcanic eruptions, but by large-scale changes to climate and environments caused by volcanism. Indirect effects resulting in respiratory distress, toxicity, famine, and habitat destruction have accounted for approximately 8 million deaths during historical times, while direct effects have accounted for fewer than 200,000, or 2.5% of the total.

12.5 Forecasting Earthquakes and Minimizing Impacts

It has long been a dream of seismologists, geologists, and public safety officials to be able to accurately predict the location, magnitude, and timing of earthquakes on time scales that would be useful for minimizing danger to the public and damage to infrastructure (e.g., weeks, days, hours). Many methods of prediction that have been explored. These include using observations of warning foreshocks, changes in magnetic fields, episodic tremor and slip, changing groundwater levels, strange animal behaviour, patterns in the timing between earthquakes, and how stress is transferred after a rupture. So far, none of these has provided a reliable method. Although there are some reports of successful earthquake predictions, they are rare, and many are surrounded by doubtful circumstances.

The Parkfield Prediction Experiment

There was great hope for earthquake predictions late in the 1980s when attention was focused on part of the San Andreas Fault at Parkfield, approximately 200 km south of San Francisco, California. Between 1881 and 1965 there were five earthquakes at Parkfield. They were spaced at approximately 20-year intervals, all confined to the same 20 km-long segment of the fault, and all very close to M6 (Figure 12.33). Both the 1934 and 1966 earthquakes were preceded by small foreshocks exactly 17 minutes before the main quake.

The U. S. Geological Survey recognized this as an excellent opportunity to understand earthquakes and earthquake prediction, so they armed the Parkfield area with a huge array of geophysical instruments and waited. The next earthquake was expected to happen around 1987, but nothing happened! The “1987 Parkfield earthquake” finally struck in September 2004. Fortunately, all of the equipment was still there to record the earthquake, but it was no help from the perspective of earthquake prediction. There were no significant precursors to the 2004 Parkfield earthquake in any of the parameters measured, including tremors, changes in rock deformation, the magnetic field, the electrical conductivity of the rock, and **creep**

(motion along the fault that is not accompanied by earthquakes). There was no foreshock. In other words, even though every available technique was used to monitor it, the 2004 earthquake came with no warning whatsoever.

Figure 12.33 Earthquakes on the Parkfield segment of the San Andreas Fault between 1881 and 2004. *Source: Steven Earle (2015) CC BY 4.0*

Earthquake Probabilities

To be useful to the public and governments, earthquake predictions must be accurate *most* of the time, not just *some* of the time. If a prediction method is only accurate 10% of the time (and even that isn't possible with the current state of knowledge), the public will lose faith in the process very quickly, and then will ignore all of the predictions. The hope for earthquake prediction is not dead, but it was hit hard by the Parkfield experiment.

Today the focus of efforts in earthquake-prone regions is to provide forecasts of **earthquake probability**. Earthquake probabilities express the likelihood that an earthquake of a given magnitude will occur at a location within a given period of time. An example of this approach for the San Francisco Bay region of California is shown in Figure 12.34. Based on a wide range of information, including past earthquake history, accumulated stress from plate movement, and known stress transfer, seismologists and geologists have predicted the likelihood of a M6.7 or greater earthquake on each of eight major faults that cut through the region. The greatest probabilities are on the San Andreas, Rogers Creek/Hayward faults, and Calaveras/Paicines faults. There is a 72% chance that a major and damaging earthquake will take place somewhere in the region prior to 2043.

Using Earthquake Probability Information

Decision makers can use forecasts of earthquake probability to assist with educating the public about earthquake risks, and to determine what action is necessary to make infrastructure earthquake-safe. Building safe infrastructure requires strong building codes, and enforcement of those codes. Building code compliance is robust in most developed countries, but is inadequate in many developing countries.

New buildings are not the only ones requiring attention. Existing buildings — especially schools and hospitals — and other structures such as bridges and dams, must also be made safe. British Columbia began a multi-billion-dollar program in 2004 to make public schools safer for students. The program is focused on older public schools, because, according to the government, those built since 1992 already comply with modern seismic codes. Some schools would require too much work to make upgrading economically feasible and they are replaced. Where upgrading is feasible, the school is assessed carefully before any upgrade work is initiated. The seismic mitigation program identified 346 schools as being at high risk and in need of upgrades. As of December 2017, upgrades were completed at 168 schools, 28 schools were under construction or had approval to proceed with construction, and 150 did not yet have plans in place for upgrades.

Figure 12.34 Earthquake outlook for 2014-2047. Probabilities for individual faults are marked on the faults. *Source: U. S. Geological Survey (2014) Public Domain*

The program in British Columbia illustrates a challenge with seismic upgrades of public buildings, which is that governments must make adequate funds available for the upgrades to be done in a timely manner. The priority allocated to funding those projects will depend on how urgent the need for upgrades is considered to be, given other demands on public funds.

Earthquake Preparedness

Earthquake preparedness involves the formulation of public emergency plans, including escape routes, medical facilities, shelters, and food and water supplies. It also includes personal planning, such as emergency supplies (food, water, shelter, and warmth), escape routes from houses and offices, and communication strategies (with a focus on ones that don't involve the cellular network).

Chapter 12 Summary

The topics covered in this chapter can be summarized as follows:

12.1 What Is an Earthquake?

An earthquake is the shaking that results when a deformed body of rock snaps back to its original shape. The rupture is initiated at a point but quickly spreads across the area of a fault, with aftershocks initiated by stress transfer. Episodic tremor and slip is a periodic slow movement, accompanied by harmonic tremors, along the middle part of a subduction zone boundary.

12.2 Measuring Earthquakes

Earthquakes produce seismic waves that can be measured by a seismograph. The amplitudes of seismic waves are used to determine the amount of energy released by an earthquake- its magnitude. For the moment magnitude scale used today, the amount of energy released by an earthquake is proportional to the size of the rupture surface, the amount of displacement, and the strength of the rock. Intensity is a measure of the amount of shaking that occurs, and damage done at locations that experience an earthquake. Intensity will vary depending on the distance to the epicentre, the depth of the earthquake, and the type of geological materials present.

12.3 Earthquakes and Plate Tectonics

Most earthquakes happen at or near plate boundaries. Along divergent and transform boundaries earthquakes are shallow (less than 30 km depth), but at convergent boundaries they can be hundreds of kilometers beneath the surface. The largest earthquakes happen when a broad segment of the locked zone of a subduction zone ruptures. Intraplate earthquakes happen away from plate boundaries. They can be caused by human activities, or renewed motion on ancient faults.

12.4 The Impacts of Earthquakes

Damage to buildings is the most serious consequence of most large earthquakes. The amount of damage is related to the type and size of buildings, how they are constructed, and the nature of the material on which they are built. Other important consequences are fires, damage to bridges and highways, slope failures, liquefaction, and tsunami.

12.5 Forecasting Earthquakes and Minimizing Impacts

There is no reliable technology for predicting earthquakes, but the probability of one happening within a certain time period can be forecast. We can minimize earthquake impacts by ensuring that the public is aware of the risk, that building codes are enforced, that existing buildings like schools and hospitals are seismically sound, and that both public and personal emergency plans are in place.

Review Questions

1. What causes the shaking during an earthquake?
2. What is a rupture surface, and how does the area of a rupture surface relate to earthquake magnitude?
3. What is an aftershock and how are aftershocks related to stress transfer?
4. Episodic slip on the middle part of the Cascadia subduction zone is thought to increase the stress on the locked zone. Why?
5. What is the difference between the magnitude of an earthquake and its intensity?
6. How much more energy is released by a magnitude 7 earthquake compared to a magnitude 5 earthquake?
7. The images below show earthquake locations for three regions of ocean lithosphere. The colour scheme indicates the depths of earthquakes. a) The images show a subduction zone, a slowly spreading mid-ocean ridge, and a rapidly spreading mid-ocean ridge. Which is which? b) In the

image with the subduction zone, which side is the subducting plate, and which is the overriding plate?

Figure 12.35
Earthquakes along plate margins. Dots are colour coded according to depth. *Source: Map details from Lisa Christiansen, Caltech Tectonics Obs. (2008). See Appendix C for terms of use.*

8. Why is earthquake damage likely to be more severe for buildings built on unconsolidated sediments as opposed to on solid rock?
9. Why are fires common during earthquakes?
10. What type of earthquake is likely to lead to a tsunami?
11. What did we learn about earthquake prediction from the 2004 Parkfield earthquake?
12. What are some of the things we should know about an area in order to help minimize the impacts of an earthquake?
13. What is the difference between earthquake prediction and forecasting earthquake probability?

Answers at the end of the chapter

References

Ammon, C. J. (2001). Earthquake Size. http://eqseis.geosc.psu.edu/~cammon/HTML/Classes/IntroQuakes/Notes/earthquake_size.html

Bao, X., and Eaton, D. W. (2016). Fault activation by hydraulic fracturing in western Canada. *Science 354*(6318), 1406-1409. doi: 10.1126/science.aag2583

Buck, W. R., Lavier, L. L., & Poliakov, A. N. B. (2005). Modes of faulting at mid-ocean ridges. *Nature 434*(7034), 719-23. doi: 10.1038/nature03358
Coastal and Marine Geology Program, U. S. Geological Survey (n.d.). Lake Baikal - A Touchstone for Global Change and Rift Studies. https://pubs.usgs.gov/fs/baikal/

DeMets, C., Gordon, R. G., & Argus, D. F. (2010). Geologically current plate motions. *Geophysical Journal International 181*, 1-80. doi: 10.1111/j.1365-246X.2009.04491.x

Earthquakes Canada, Natural Resources Canada (2016) Earthquake zones in Eastern Canada http://www.seismescanada.rncan.gc.ca/zones/eastcan-en.php

Earthquakes Canada, Natural Resources Canada (2016). The M9 Cascadia Megathrust Earthquake of January 26, 1700. http://www.earthquakescanada.nrcan.gc.ca/historic-historique/events/17000126-en.php

Finn, W. D. L., & Dexter, A. (2012). Risk Management Plan for School Seismic Mitigation Program. https://www2.gov.bc.ca/assets/gov/education/administration/resource-management/capital-planning/risk_management_plan_school_seismic_mitigation_program_2012.pdf

Gendzwill, D. J., Horner, R. B., & Hasegawa, H. S. (1982). Induced earthquakes at a potash mine near Saskatoon, Canada. *Canadian Journal of Earth Science 19*, 466-475. doi: 10.1139/e82-038

Green, H. W. II, Chen, W.-P., & Brudzinski, M. R. (2010). Seismic evidence of negligible water carried below 400-km depth in subducting lithosphere. *Nature 467*, 828-830. doi:10.1038/nature09401

Hongyu, Y., Liu, Y., Harrington, R. M., & Lamontagne, M. (2016). Seismicity along St. Lawrence Paleorift Faults Overprinted by a Meteorite Impact Structure in Charlevoix, Québec, Eastern Canada. *Bulletin of the Seismological Society of America 106*(6), 2663-2673.
doi: https://doi.org/10.1785/0120160036

Ives, J. D., Messerli, B. (1989). *The Himalayan Dilemma: Reconciling development and conservation.* New York: Routledge.

Milne, W. G. (1955). Seismology in British Columbia. *The Journal of the Royal Astronomical Society of Canada XLIX*(4), 141 - 150. http://openpress.usask.ca/physicalgeology/wp-content/uploads/sites/29/2017/11/1955JRASC__49__141M.pdf

Ministry for Culture and Heritage (2017). Christchurch earthquake kills 185.

https://nzhistory.govt.nz/page/christchurch-earthquake-kills-185

Myhill, R., & Warren, L. M. (2012). Fault plane orientations of deep earthquakes in the Izu-Bonin-Marianas subduction zone. *Journal of Geophysical Research 117,* B06307. doi:10.1029/2011JB009047

Nishikawa, T., & Ide, S. (2015). Background seismicity rate at subduction zones linked to slab-bending-related hydration. *Geophysical Research Letters 42,* 7081-7089. doi: 10.1002/2015GL064578

Province of British Columbia (n.d.). Seismic Mitigation Program. https://www2.gov.bc.ca/gov/content/education-training/administration/resource-management/capital-planning/seismic-mitigation-program

Province of British Columbia (2017). Seismic Mitigation Program Progress Report. https://www2.gov.bc.ca/assets/gov/education/administration/resource-management/capital-planning/seismic-mitigation/progress_report.pdf

Sloan, R. A., Jackson, J. A., McKenzie, D., & Priestley, K. (2011). Earthquake depth distributions in central Asia, and their relations with lithosphere thickness, shortening and extension. *Geophysics Journal International 185,* 1-29. doi: 10.1111/j.1365-246X.2010.04882.x

University Corporation for Atmospheric Research (2010). Propagation. *Tsunamis.* http://www.torbenespersen.dk/Publish/tsunami/print_4.htm#page_1.1.2

U. S. Geological Survey (2014). Indian Ocean Tsunami Remembered — Scientists reflect on the 2004 Indian Ocean that killed thousands. https://www2.usgs.gov/blogs/features/usgs_top_story/indian-ocean-tsunami-remembered-scientists-reflect-on-the-2004-indian-ocean-that-killed-thousands/

Wang, K., Jiangheng, H., Schulzeck, F., Hyndman, R. D., and Riedel, M. (2015). Thermal Condition of the 27 October 2012 Mw 7.8 Haida Gwaii Subduction Earthquake at the Obliquely Convergent Queen Charlotte Margin. *Bulletin of the Seismological Society of America, 105*(2B), 1290–1300. doi: 10.1785/0120140183

Additional Resources

Lamontagne, M., Halchuk, S., Cassidy, J. F., & Rogers, G. C (2008) Significant Canadian Earthquakes of the Period 1600 - 2008. *Seismological Research Letters 79*(2), 211 - 223. doi: 10.1785/gssrl.79.2.211 http://openpress.usask.ca/physicalgeology/wp-content/uploads/sites/29/2017/11/SRL2008_Lamontagne_etal.pdf

Detailed description of the effects of the 1946 earthquake: Hodgson, E. A. (1946). British Columbia Earthquake, June 23, 1946. *The Journal of the Royal Astronomical Society of Canada XL*(8), 285 - 319. http://openpress.usask.ca/physicalgeology/wp-content/uploads/sites/29/2017/11/1946JRASC__40__285H.pdf

Latest Earthquakes (U. S. Geological Survey) https://earthquake.usgs.gov/earthquakes/map/

IRIS Teachable Moment slides for the October 2012 Haida Gwaii earthquake: http://openpress.usask.ca/physicalgeology/wp-content/uploads/sites/29/2017/11/121028queencharlotte.pdf

Rogers, G. & Dragert, H. (2003). Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip. *Science, 300,* 1942-1943.

Answers to Review Questions

1. Rocks under stress can deform elastically. When rocks break, or a rupture occurs along an existing fault, the deformed rocks snap back to their original shape, causing vibrations. The vibrations are the shaking felt during earthquakes.

2. The rupture surface is the surface over which there is displacement of rock during an earthquake. The magnitude of an earthquake is proportional to the area of the rupture surface.

3. An aftershock is an earthquake that is triggered when a rupture from a previous earthquake has transferred too much stress to rocks, causing further ruptures.

4. Episodic slip on the middle part of the Cascadia subduction zone decreases stress within that area, but some of that stress is transferred to the locked zone along the plate boundary. This increases the amount of stress on the locked part.

5. Magnitude is the amount of energy released by an earthquake. Each earthquake has only one magnitude, although there are different ways of measuring it, and they may give slightly different results. Intensity is a measure of the amount of damage done or what people felt. Intensity varies depending on the distance to the epicentre and the type of rock or sediment underlying an area.

6. Each unit increase in magnitude corresponds to a 32x increase in energy. The difference between M5 and M7 is two units of magnitude. An earthquake that is two magnitude units larger would release 32 x 32 = 1,024 times as much energy.

7. a) A: Slowly spreading mid-ocean ridge- earthquakes are shallow and evenly distributed; B: Subduction zone- earthquakes range in depth; C: Rapidly-spreading mid-ocean ridge- earthquakes are shallow, and occur in patches, corresponding to a greater likelihood of earthquakes where transform faults offset ridges.

b) The plate to the right of the chain of earthquakes is the subducting plate. The earthquakes increase in depth moving from right to left, indicating that the subducting slab is getting deeper farther to the left.

8. Unconsolidated sediments, especially if they are saturated with water, can lose strength when subjected to earthquake shaking. This can cause buildings to subside or tilt. Unconsolidated sediments can also amplify the vibrations of an earthquake.

9. Gas lines and electrical transmission wires are typically damaged during an earthquake, and this can lead to serious fires.

10. A large subduction-zone earthquake (greater than M7.5) can generate a tsunami because those earthquakes result in sufficient vertical displacement of the sea floor.

11. The 2004 Parkfield earthquake showed that we cannot rely on foreshocks to predict earthquakes, or on any of the many other parameters that were being carefully measured around Parkfield in the years leading up to the quake.

12. We should know about the history of past large earthquakes, the typical locations of small earthquakes, the types of geological materials beneath the surface (especially soft water-saturated sediments), the types of infrastructure that is present, and the various ways that people can be evacuated from an area or assistance can be brought in.

13. Forecasting involves estimating the risk of an earthquake happening in a region within a period of time. Prediction involves stating that an earthquake is likely to happen at a certain location on a specific day or month or year in the future. With our current state of knowledge of earthquakes, prediction is not possible.

Physical Geology, 3rd Adapted Edition is used under a CC BY 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 13. Geological Structures & Mountain Building

by Karla Panchuk, University of Saskatchewan

Figure 13.1 Folded rocks in the Cariboo Mountains of BC. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photograph by Drew Brayshaw (2009) CC BY-NC 2.0. See Appendix C for more attributions.*

Learning Objectives

After reading this chapter and answering the *Review Questions* at the end, you should be able to:

- Describe the types of stresses that affect rocks.
- Explain how rocks respond to those stresses through brittle, elastic, or plastic deformation.
- Explain how rocks become folded and know the terms used to describe fold characteristics.
- Describe the conditions under which rocks fracture.
- Describe the different types of faults (normal, reverse, thrust, strike-slip) and the stresses that create them.
- Describe the different ways in which deformation of Earth's lithosphere builds mountains.
- Measure the strike and dip of a geological feature, and plot the information on a map.

Folds, like those in the centre of Figure 13.1, are a common feature of mountain belts. Have you ever wondered how something as hard as rock could flex and bend to make folds, and what forces are required? Geologists have.

Observing and analysing geological structures helps us to understand the kinds of forces that affect rocks, both on small scales, and on scales as large as tectonic plates. With that knowledge we can understand how plate tectonic processes change the shape of Earth's lithosphere. We can also piece together a history of past changes, including how mountain belts are formed. Structural geologists make careful observations of the orientations of breaks and bends in rocks, and can compile those measurements into maps of geological structures. These maps can be valuable tools for finding mineral resources.

13.1 Stress and Strain

Plate collisions and the accumulated weight of overlying rocks exert forces on rocks at depth. While the size of the force is important, it also matters whether the force is distributed over a wide region, or tightly focused on a small area. The same force will have a greater effect when acting over a small area than when acting over a larger area. If you have ever used snowshoes to walk across a snow bank without sinking in, you have taken advantage of the effects of distributing force (your mass acted upon by gravity) over a wider area (the area of your snowshoes rather than the soles of your boots). **Stress** is force adjusted for the area over which it is distributed. **Strain** is the change in shape that happens when rocks are deformed by stress.

Types of Stress

Stresses fall into two categories: **normal stress** acts at right angles to a surface, and **shear stress** acts parallel to a surface (Figure 13.2). Normal stress is subdivided into **compression**, when the stresses are squeezing a rock, and **tension**, when stress is pulling it apart. Rocks undergo compression in regions where plates are colliding, or where they are being buried beneath other rocks. Rocks experience tension where divergence is happening, such as when a continent is beginning the rifting process. Shear stress is characteristic of transform plate boundaries, where plates are moving side by side.

Figure 13.2 Rocks can be affected by normal stress (compression and tension) or shear stress. *Source: Karla Panchuk (2016) CC BY 4.0*

Although Figure 13.2 shows only one set of stress arrows for each scenario, rocks within the Earth are subject to stress from all directions. The relative size of the stresses in different directions will determine the response of the rock. Consider a deeply buried rock being stretched as a continent breaks apart (Figure 13.3). It is also being compressed by the weight of overlying sediments and rocks, but the stress from compression is relatively small compared to the tension from rifting. The net effect of stress acting on the rock will be determined more by the tension from rifting than by the compression from overlying rocks.

Figure 13.3 A rock layer (dark brown) is compressed by the weight of rocks above, and stretched by rifting. The sizes of the arrows indicate the relative sizes of the stresses. *Source: Karla Panchuk (2018) CC BY 4.0*

Rocks experience stress from all directions, but it is possible to break down stresses into three directions, just like a graph with x, y, and z axes. In diagrams showing these three directions, the sizes of arrows representing each direction will indicate the relative size of stresses, as they do in Figure 13.3. Analyzing stress in this way makes it much easier to describe the stresses operating on a rock, and to understand what their net effect will be.

Types of Strain

How a rock responds to stress depends on many factors. The "how" is not simply a matter of *how much* strain a rock will undergo, but what *type* of strain will occur. Is the deformation permanent or temporary? Does the rock break or does it deform without breaking?

Elastic Strain

Elastic strain is reversible strain. You can think of elastic strain as what happens to the elastic waistband of your favourite sweatpants when you put them on. The elastic stretches to let you into your pants, and once you're in them, it shrinks to keep them from falling down. When you take the pants off again, the elastic goes back to its original shape. Similarly, rocks undergoing elastic strain will snap back to their original shape once the stress is removed. Rocks snapping back to their original shape undergo **elastic rebound**. Elastic rebound of rocks on a large scale can have profound consequences, because the energy released causes the Earth to vibrate. We experience those vibrations as earthquakes.

Plastic Strain

If enough stress is applied, the changes that a material undergoes to accommodate the stress will leave it permanently deformed. When the stress is removed, the material does not go back to its original shape. The permanent deformation is called **plastic strain**.

Ductile or Brittle?

Ductile deformation refers to deformation that happens by flowing or stretching. The marble monument in Figure 13.4 is undergoing ductile deformation as it sags beneath its own weight.

When a material breaks, it has undergone **brittle deformation** (Figure 13.5). The stone cylinders in Figure 13.5 are part of an experiment to test the strength of the rock. The cylinder on the right looked like the cylinder on the left before it was compressed, with force applied to the top and bottom. Strain gauges have been glued on to measure the amount of deformation lengthwise and across the cylinders.

Figure 13.4 A marble monument at the Westminster Hall and Burying Ground in Baltimore, Maryland. The horizontal surface is undergoing slow ductile deformation as it sags beneath its own weight. The author Edgar Allan Poe is interred nearby. *Source: Ray Pennisi (2007) CC BY-NC 2.0*

Figure 13.5 Cylinders of rock used to test the strength of rock under compression. The cylinder on the left has been equipped with strain gauges to measure the amount of deformation. The cylinder on the right has undergone brittle deformation after being compressed. *Source: Karla Panchuk (2016) CC BY 4.0*

A material can undergo more than one kind of deformation when stress is applied. The barrel-shaped cylinder of potash in Figure 13.6 (right) originally looked like the cylinder on the left. The cylinder was compressed, with stress applied from the top and bottom. Initially, it underwent ductile deformation and thickened in the middle, creating the barrel shape. But as more stress was applied, the cylinder eventually underwent brittle deformation, resulting in the crack across the middle.

Figure 13.6 Cylinders of potash before and after deformation. The potash underwent ductile deformation before it finally broke. *Source: Karla Panchuk (2018) CC BY 4.0*

Factors That Determine How A Rock Will Deform

A rock is not limited to exclusively brittle deformation, or exclusively ductile deformation. Even the deformed rock in Figure 13.5, which has clearly undergone brittle deformation, shows a slight curvature on the right side, near the top. This indicates that a small amount of ductile deformation occurred before brittle failure.

For a given rock, deformation will be different depending on the amount of stress applied. Up to a point, rocks undergo elastic deformation, and will spring back to their original shape after the stress is removed. If more stress is applied, the rock may deform in a ductile manner. If stress increases further, the rock may fracture. The amount of stress required in each case will depend on the type of rock, as well as conditions such as pressure and temperature.

Composition

In general, sedimentary rocks will be more likely to undergo ductile deformation than igneous or metamorphic rocks under the same conditions. Rocks within each group will also deform differently.

Boudinage structures (Figure 13.7) highlight the effect of composition on how rocks deform. These structures occur when a stronger rock more prone to brittle deformation is surrounded by weaker rocks prone to ductile deformation. The stronger rock will fracture into segments, called **boudins**, and the weaker rock will flow into the spaces between. In Figure 13.7 (top), the white layer reached the stage of pinching off, just before separating into segments. The surrounding black layer flowed in to fill the gap where the pinch was happening. Remarkably, the white layer itself contains a dark layer that has fragmented into boudins. Not all boudins break into blocky segments. Some display more ductile deformation (Figure 13.7, bottom).

Figure 13.7 Two examples of boudinage structure. Top- The white layer has pinched off into segments, and the surrounding black layers have flowed into the gap forming between segments. Within the white layer is a thinner black layer that has also broken into segments. Bottom- Boudins displaying ductile deformation. *Source: Top- Marek Cichanski (2012) CC BY-NC 2.0. Bottom- Joyce McBeth (n.d.) CC BY 4.0*

Temperature and Pressure

At higher temperatures, and under higher confining pressures, rocks are more likely to undergo ductile deformation. **Confining pressure** is the stress that a material experiences uniformly from all sides as a result of the weight of material above and around it. The pressure that a diver feels deep in the ocean is confining pressure due to the weight of water above and around the diver. This kind of confining pressure

is called **hydrostatic pressure**. Within Earth, the confining pressure is due to the weight of overlying rocks. Confining pressure due to the weight of rocks is called **lithostatic pressure.**

The rocks in Figures 13.5 and 13.6 experienced confining pressure from the atmosphere, and temperatures comfortable for the humans working in the lab. Under those conditions the rocks ultimately underwent brittle failure when they were compressed in the lab. Deep within the crust, the temperatures and confining pressures are far greater. Deep enough within the crust, both samples would undergo only ductile deformation if the same amount of stress were applied as in the experiment. The depth at which temperatures and confining pressures are high enough for rocks to go from brittle deformation to ductile deformation is called the **brittle-ductile transition zone**.

The brittle-ductile transition zone occurs between approximately 10 km and 30 km depth, corresponding to temperatures around 300 ºC and greater. The depth at which temperatures reach 300 ºC at any particular location will depend on heat flow at that location. In continental crust, rocks at 300 ºC are deeper than in ocean crust. The change in pressure with depth also varies, depending on the mass and density of rocks. If depths are measured relative to sea level, the pressure at 10 km measured beneath a tall mountain belt will be greater than the pressure at 10 km measured within ocean crust.

Experiments like those shown in Figures 13.5 and 13.6 can be used to determine where the brittle-ductile transition zone will be for a particular rock type. Experimenters apply stress to sample of a rock for a range of temperatures and confining pressures. They note the conditions under which the rock breaks or deforms in a ductile manner, and plot those on a graph (Figure 13.8). The results in Figure 13.8 are from experiments on limestone. The vertical axis is pressure. The more pressure, the deeper the rock would have to be within the Earth to experience that pressure. The white line represents the brittle-ductile transition zone. Above the white line are pressures and temperatures under which the limestone would fracture. Below the white line in the tan area are pressures and temperatures where the limestone would deform by flowing. Notice that the higher the temperatures, the less confining pressure is required for ductile deformation.

Figure 13.8 Experimental results on limestone with tension applied (left) and compression applied (right). *Source: Karla Panchuk (2018) CC BY 4.0 modified after Heard (1960)*

How Stress Is Applied

The limestone experiments were performed by applying stress as tension (Figure 13.8 left) and again by applying stress as compression (right). When tension was applied, temperature and confining pressure had to be much higher before ductile deformation occurred. Under compression, ductile deformation was possible with far less confining pressure, and at lower temperatures.

Strain rate, the rate at which deformation occurs, also makes a difference. If stress is applied at a rate that causes rapid deformation, the rock will be more likely to fracture than if deformation happens slowly. The marble slab in Figure 13.4 is a good example of this. It has sagged rather than broken because the rate of deformation has been very slow, at millimetres per decade.

Fluids

When rocks are under pressure, fluids trapped within the **pore spaces** of rocks- the gaps between grains- are also under pressure. Higher confining pressure is required for deformation to be ductile rather than brittle, but pressure from fluids, called **pore pressure**, resists the confining pressure. The result is that the effective confining pressure is lower than it would be without the fluids. Depending on the amount of pore pressure, and how close the rock is to the brittle-ductile transition zone, pore pressure could cause brittle failure in a rock that would otherwise undergo ductile deformation.

Stress and Geological Structures

Many different geologic structures can form when stress is applied to rocks. Structures form as a result of fracturing, tilting, folding, stretching, and squeezing (Figure 13.9). Some structures, like the fractures that make basalt columns (Figure 13.9, upper left), happen when rocks shrink due to cooling, but others are a consequence of plate tectonic forces. The types of structures that form depend on the plate tectonic setting and other geological conditions, making them valuable tools for understanding what happened to the rocks. The following sections address the different kinds of structures that form, and what information we can gather from these structures to learn more about the tectonic environment and regional geology.

Figure 13.9 Structures resulting from deformation. Upper left- Fracturing in basalt near Whistler, British Columbia. Upper right- Tilting of sedimentary rock near Exshaw, Alberta. Lower left- Stretched limestone (light grey) and chert (dark grey) from Quadra Island, British Columbia. Lower right- Faulted shale near Cache Creek, British Columbia. Rocks above the fault moved up relative to those below. *Source: Karla Panchuk (2018) CC BY 4.0. Photographs by Steven Leslie (2015) CC BY 4.0*

13.2 Folds

Folds are a type of ductile deformation. They form when rocks bend in response to stress. The sides of a fold are its **limbs** (Figure 13.10). The limbs meet in a region of curvature called the **hinge zone**. A fold's **axial surface** is an imaginary surface that runs along the hinge zone and cuts the fold in half. The line that forms when the axial surface intersects another surface, such as the top of a bed, is called the **axial trace**. Axial traces are sometimes marked on geological maps to show the location of the fold's hinge zone.

Figure 13.10 The parts of a fold. A fold consists of limbs that meet at the hinge zone. An axial surface bisects the fold along the hinge zone. The axial trace is where the axial surface intersects another surface, such as the top of a bed. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photo: Ron Schott (2009) CC BY-NC-SA 2.0*

Fold Classification

Synclines and Anticlines

Folds can be classified according to the whether the limbs slope toward or away from the hinge zone. If the limbs slope toward the hinge zone (i.e., the hinge zone points downward), as in the fold in the left of Figure 13.11, the fold is called a **syncline**. If the limbs slope away from the hinge zone (i.e., the hinge zone points upward), the fold is called an **anticline**. There is an anticline on the right side of Figure 13.11. The fold in Figure 13.10 is also an anticline. Sometimes an anticline or a syncline will occur by itself, but they can also occur in a series of alternating synclines and anticlines, similar to the way the anticline and syncline share a limb in Figure 13.11. A sequence of linked anticlines and synclines is called a **fold train**.

Symmetrical, Asymmetrical, Overturned, and Recumbent

In a **symmetrical fold**, the limbs slope at approximately the same angle on either side of the axial surface. The fold in Figure 13.10 is symmetrical. In an **asymmetrical fold**, the limbs slope at different angles on either side of the axial surface. The syncline in Figure 13.11 is asymmetrical. The limb on the left side of the syncline slopes toward the hinge at a steeper angle than the limb on the right.

If the fold is sufficiently tilted that the beds on one side have been tilted past vertical, and are sloping in the same direction, the fold is **overturned** (Figure 13.12).

Figure 13.11 An asymmetrical syncline linked to an anticline on a beach in Cornwall, United Kingdom. The beds slope toward the hinge at different angles on either side of the axial surface. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photo: Harry Soar (2014) CC BY-NC-SA 2.0*

It is possible for rocks to be folded so tightly that the fold limbs are nearly parallel. Folds with parallel limbs are called **isoclinal folds**. A **recumbent fold** is an isoclinal fold that has been overturned to the extent that the limbs are horizontal (Figure 13.13).

Figure 13.12 Overturned folds in Andalusia in southern Spain. Some limbs have been overturned far enough to be sloping in the same direction on either side of the axial trace. *Source: Karla Panchuk (2018) CC BY-NC-SA 2.0. Photo: Ignacio Benvenuty Cabral (2017) CC BY-NC-SA 4.0*

Figure 13.13 A recumbent fold has limbs that are nearly parallel, and an axial trace that is nearly horizontal. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Photo: Ignacio Benvenuty Cabral (2017) CC BY-NC-SA 4.0*

Folds in the Landscape

Folds can be of any size, and it's very common to have smaller folds within larger folds (Figure 13.14). Large folds can extend over 10s of kilometres, and very small ones might only be visible under a microscope.

Figure 13.14 Folded limestone (grey) and chert (rust-coloured) in rocks of the Triassic Quatsino Formation on Quadra Island, British Columbia. The image is about 1 m across. *Source: Steven Earle (2015) CC BY 4.0*

When folded rocks are weathered and eroded, they can alter the landscape by forming long ridges and valleys (Figure 13.15). Ridges and valleys curve into V-shapes if the hinge of the fold is not horizontal. A fold with a hinge that slopes downward is called a **plunging fold** (Figure 13.16).

Figure 13.15 Ridges and valleys in central Pennsylvania formed from weathered and eroded folds. The V-shapes indicate the folds are plunging. *Source: NASA on the Commons (2001) Public Domain*

Figure 13.16 Plunging folds have sloping hinges. Plunging folds are described in terms of the plunge angle, the angle the hinge makes with a horizontal line. Inset- When a plunging fold intersects a surface, the result is a V-shaped pattern. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Photo- Dieter Mueller (2004) CC BY-SA 3.0*

Folds can create landforms, but anticlines are not necessarily expressed as ridges in the terrain. Likewise, synclines do not necessarily appear as valleys. When folded rocks erode, the landform that results depends how resistant individual layers are to erosion. For example, if the rocks in the interior of an anticline are more resistant to weathering than the surrounding rocks, a ridge will result (e.g., the low hill represented by units 4 and 5 in Figure 13.17, top). On the other hand, if rocks in the interior of the anticline are weaker, a valley will result (Figure 13.17, bottom, units d^1 and d^2). Similarly, a syncline with stronger rocks in the interior will weather to form a ridge, and a syncline with weaker rocks in the interior will weather to form a valley.

Figure 13.17 Cross-sections of eroded folds expressed as hills and valleys, from an early study on the geology of Wales, Devon, and Cornwall. Top- An anticline in Shropshire, England. Beds in the interior of the anticline form a gentle hill. Bottom- An anticline in Herefordshire, England in which beds in the interior of an anticline weathered to form a valley. *Source: Symonds (1872) Public Domain.*

Exercise: Fold Types

What kind of folds are shown here?

Figure 13.18 Folds in the Rocky Mountains near Golden, British Columbia. *Source: Steven Earle (2015) CC BY 4.0*

13.3 Fractures, Joints, and Faults

When rocks break in response to stress, the resulting break is called a **fracture**. If rocks on one side of the break shift relative to rocks on the other side, then the fracture is a **fault**. If there is no movement of one side relative to the other, and if there are many other fractures with the same orientation, then the fractures are called **joints**. Joints with a common orientation make up a **joint set** (Figure 13.19).

Figure 13.19 Joint sets have broken these siltstone and shale beds into long rectangular planks. *Source: Michael C. Rygel (2008) CC BY-SA 3.0*

Jointing

Most joints form when the overall stress regime is one of tension (pulling apart) rather than compression. The tension can be from a rock contracting, such as during the cooling of volcanic rock (Figure 13.9, upper left). It can also be from a body of rock expanding. **Exfoliation joints**, which make the rock appear to be flaking off in sheets (Figure 13.20), occur when a body of rock expands in response to reduced pressure, such as when overlying rocks have been removed by erosion.

Figure 13.20 Half Dome at Yosemite National Park is an exposed granite batholith that displays exfoliation joints, causing sheets of rock to break off. *Source: HylgeriaK (2010) CC BY-SA 3.0*

Nevertheless, it is possible for joints to develop where the overall regime is one of compression. Joints can develop where rocks are being folded, because the hinge zone of the fold is under tension as it stretches to accommodate the bending (Figure 13.21). Joints can also develop in a rock a rock under compression as a way to accommodate the change in shape (Figure 13.22). The joints accommodate the larger compression stress (larger red arrows) by allowing the rock to stretch in the up-down direction (along the green arrows).

Figure 13.21 Joints developed in the hinge zone of folded rocks. *Source: Steven Earle (2015) CC BY 4.0*

Figure 13.22 Joints developing to accommodate the larger horizontal component of compression (large red arrows). *Source: Steven Earle CC BY 4.0*

Faulting

A fault is a boundary between two bodies of rock along which there has been relative motion (e.g., Figure 13.23). Some large faults, like the San Andreas fault in California or the Tintina fault, extending from northern British Columbia through central Yukon and into Alaska, show evidence of hundreds of kilometres of motion. Other faults show only centimetres of movement. In order to estimate the amount of motion on a fault, it is necessary to find a feature that shows up on both sides of the fault, and has been offset by the fault. This could be the edge of a bed or dike as in Figure 13.23, or it could be a landscape feature, such as a fence or a stream.

Figure 13.23 View looking down on a fault (white dashed line) in intrusive rocks on Quadra Island, British Columbia. The pink dyke has been offset approximately 10 cm by the fault (length of the white arrow). *Source: Steven Earle (2015) CC BY 4.0*

Types of Faults

Different kinds of faults develop under different stress conditions. We describe faults in terms of how the rocks on one side of the fault move relative to the other.

Dip-Slip Faults

Dip-slip faults are so named because the dominant motion involves moving up or down the dipping (tilting) fault plane. In dip-slip faults we identify rock above the fault as the **hanging wall**, (or **headwall**) and the rock beneath as the **footwall**. These terms were originally used by miners to describe the rocks above and below an ore body (Figure 13.24).

Figure 13.24 The hanging wall (or headwall) of a fault is the rock above the fault. The footwall is the rock below. These terms were originally used by miners to describe the rocks above and below an ore body. *Source: Photo- Gold Hill Mine, Yukon Territory, by Eric A. Hegg (1898) Public Domain. Diagram- Karla Panchuk (2018) CC BY 4.0*

Tension produces **normal faults**, in which the crust undergoes extension. This permits the hanging wall to slide down the footwall in response to gravity (Figure 13.25, left). Compression produces **reverse faults**, pushing the hanging wall up relative to the footwall. Reverse faults shorten and thicken the crust (Figure 13.25, right).

Figure 13.25 Dip slip faults. Normal faults are caused by tension, while reverse faults happen during compression. *Source: Karla Panchuk (2018) CC BY-SA 4.0, modifed after Woudloper (2010) CC BY-SA 3.0*

Strike-Slip Faults

Faults where the motion is mostly horizontal and along the "strike" or the length of the fault are called **strike-slip faults** (Figure 13.26 bottom). These happen where shear stress causes bodies of rock to slide sideways with respect to each other, as is the case along a transform boundary. If the far side moves to the right, as in Figures 13.23 and 13.26 (right), it is a **right-handed** or **dextral** strike-slip fault. If the far side moves to the left it is a **left-handed** or **sinistral** strike-slip fault.

Figure 13.25 Dip slip faults. Normal faults are caused by tension, while reverse faults happen during compression. *Source: Karla Panchuk (2018) CC BY-SA 4.0, modifed after Woudloper (2010) CC BY-SA 3.0*

Different Tectonic Settings Have Distinct Types of Faults

Horst and Graben Structure

In areas that are characterized by extensional tectonics, and with many normal faults arranged side-by-side, some blocks may subside (settle downward) relative to neighbouring parts. This is typical in areas of continental rifting, such as the Great Rift Valley of East Africa or in parts of Iceland. In such situations, blocks that move down relative to the other blocks are **graben**, and elevated blocks with graben on either side are called **horsts**. There are many horsts and grabens in the Basin and Range area of the western United States, especially in Nevada. Part of the Fraser Valley region of British Columbia, in the area around Sumas Prairie, is a graben.

Figure 13.27 Graben and horst structures form where extension is happening. All of the faults are normal faults. *Source: Steven Earle (2015) CC BY 4.0*

Thrust Faults

Thrust faults are a type of reverse fault with a very low-angle fault plane. The fault planes of thrust faults typically slope at less than 30°. Thrust faults are relatively common in mountain belts that were created by continent-continent collisions. Some represent tens of kilometres of thrusting, where thick sheets of sedimentary rock have been pushed up and over other layers of rock (Figure 13.28).

Figure 13.28 A thrust fault. Top: prior to faulting. Bottom: after significant fault offset. *Source: Steven Earle (2015) CC BY 4.0*

There are numerous thrust faults in the Rocky Mountains, and a well-known example is the McConnell Thrust, along which a sequence of sedimentary rocks about 800 m thick has been pushed for about 40 km from west to east over underlying rock (Figure 13.29). The thrusted rocks range in age from Cambrian to Cretaceous, so in the area around Mt. Yamnuska Cambrian-aged rock (around 500 Ma) has been thrust over, and now lies on top of Cretaceous-aged rock (around 75 Ma) (Figure 13.30).

Figure 13.29 The McConnell Thrust in the eastern part of the Rockies. The rock within the faded area has been eroded. *Source: Steven Earle (2015) CC BY 4.0*

Figure 13.30 The McConnell Thrust at Mt. Yamnuska near Exshaw, Alberta. Cambrian limestones have been thrust over top of Cretaceous mudstone. *Source: Steven Earle (2015) CC BY 4.0*

Exercise: Fault Types

1. What kind of dip-slip fault is shown here?
2. What kind of stress created this fault?

Figure 13.31 A dip-slip fault. *Source: Steven Earle (2015) CC BY 4.0*

13.4 Mountain Building

Some of Earth's mountains are entirely or almost entirely the result of volcanic activity. These include volcanic islands like the Hawai'ian hotspot volcanoes, and newly formed volcanic island arcs along subduction zones. But the majority of mountain building on Earth is the result of plate tectonic forces that deform Earth's crust through faulting and folding. Mountain building through folding and faulting may or may not be supplemented by volcanic activity.

Mountain Building Along Convergent Margins

Mountain building along convergent margins is referred to as **orogeny**, and the mountains that are built are called **orogens**.

Ocean-Continent Collision

In ocean-continent collision zones, folding and faulting of rocks combines with volcanism to build mountains. An example of mountains built this way is the Sierra Nevada mountain range in Utah and Nevada. The orogeny that formed the Sierra Nevada range began around 140 million years ago.

The mountain range was built up by igneous intrusions and volcanic eruptions along a continental volcanic arc (Figure 13.32). The terrain was altered further inland as well. Sheets of rock were thrust on top of each other, and pushed inland along a detachment fault, similar to the example of the McConnell Thrust in Figure 13.29.

Figure 13.32 Orogeny in an ocean-continent collision zone. Mountains form from subduction zone volcanism, and from large sheets of rock that are thrust inland and folded. Materials accumulating on the leading edge of the continent in an accretionary wedge are eventually smashed onto the continent, adding to continental crust. *Source: Karla Panchuk (2018) CC BY 4.0. Modified after Ron Blakey, NAU Geology (n.d.). See Appendix C for terms of use.*

Continental crust flexed downward because of the weight of the mountains, and this formed a **fore arc basin** seaward of the new mountain range. Sediments accumulated within that basin. The leading edge of the continent also collected sediments and igneous rock scraped off the subducting plate, forming an **accretionary wedge**. Over time, the force of the collision would smash the basin sediments and the accretionary wedge against the continent, turning it into new continental crust.

Continent-Continent Collision

When two continents collide, it means the closure of a subduction zone, and an end to volcanism. The Alleghenian Orogeny, which brought together North America and Africa, helping to form Pangea, is an example of mountain building in a continent-continent collision zone. Before the continents came into contact with each other, mountain building on the eastern coast of North America would have involved deformation from an ocean-continent collision, as with Figure 13.32. But as subduction proceeded, the subducting plate drew Africa closer and closer to North America. The gap between the two continents began to close, and fill with sediments (Figure 13.33, top).

While a subduction zone existed, the addition of water to the mantle permitted partial melting of mantle rocks, and thus volcanic activity. However, when the two continents collided, the subduction zone closed off, and volcanism was no longer possible. As the continents smashed together, deep faults formed and stacked blocks of crust on top of each other. Old faults were reactivated. Rocks also began to shift along the boundary between an earlier orogen, the Taconic Orogen, and North America (Figure 13.33, bottom). When the continents had finally merged, Africa met North America along a **suture zone** with remnants of a continental volcanic arc on one side, and folded and faulted sedimentary rocks on the other.

Figure 13.33 Orogeny by continent-continent collision. The formation of Pangea included the merging of Africa and North America. This closed an ocean basin and stopped subduction along the coast of North America. Volcanism ended with the closure of the ocean basin, but mountains continued to grow through folding and faulting. *Karla Panchuk (2018) CC BY 4.0. Modified after Ron Blakey, NAU Geology (n.d.). See Appendix C for terms of use.*

Mountain Building Along Divergent Margins

When continents begin to split apart, normal faults form. This can lead to large blocks of crust that are tilted, raised, or lowered compared to adjacent blocks. Blocks that are elevated compared to adjacent blocks can form another type of mountain, called a **fault-block mountain**. Fault-block mountains formed in eastern North America when Pangea began to split up, and Africa pulled away from North America (Figure 13.34).

Figure 13.34 Fault-block mountains formed in a rift zone. Magma can move up along normal faults, resulting in igneous intrusions, or volcanic eruptions. Over time, valleys between elevated blocks will fill with sediment as the blocks erode. *Source: Karla Panchuk (2018) CC BY 4.0. Modified after Ron Blakey, NAU Geology (n.d.). See Appendix C for terms of use.*

Over time, elevated blocks erode, filling up valleys with sediment. The thinning of continental crust that occurs with rifting can decrease the pressure on mantle rocks enough to trigger partial melting. Magma can move up along the normal faults, forming igneous intrusions, or feeding volcanoes. The Palisades Sill in

New York and New Jersey is a result of rift-zone magmatism. It is a cliff-like feature resulting from erosion that exposed the tip of a structure like the sills in Figure 13.34.

13.5 Measuring Geological Structures

Documenting the characteristics of geological structures is used to understand the geological history of a region. One of the key features to measure is the orientation, or **attitude**, of bedding. We know that sedimentary beds are deposited in horizontal layers, so if the layers are no longer horizontal, then we can infer that tectonic forces have folded or tilted them.

The orientation of a planar feature, such as a bed of sedimentary rock, can be described with two values. The **strike** of the bed is the compass orientation of a horizontal line on the surface of the bed. The **dip** is the angle at which the surface tilts down from the horizontal (Figure 13.35). The dip is measured perpendicular to strike, otherwise the dip angle that is measured will be smaller than the actual tilt of the bed.

Figure 13.35 Strike and dip for tilted sedimentary beds. Water provides a horizontal surface. The strike and dip symbol is a T with the long horizontal bar representing the strike direction, and the small tick mark indicating the dip direction. The dip angle is written next to the tick mark. *Source: Karla Panchuk (2018) CC BY 4.0. Modified after Steven Earle (2015) CC BY 4.0*

It may help to imagine a vertical surface, such as a wall in your house. The strike is the compass orientation of the wall and the dip is 90˚ from horizontal. If you could push the wall so it is leaning over, but still attached to the floor, the strike direction would be the same, but the dip angle would be less than 90˚. If you pushed the wall over completely so it was lying on the floor, it would no longer have a strike direction because you could draw a horizontal line in any of an infinite number of directions on the horizontal surface of the wall. Its dip would be 0˚.

When reporting the dip, include the direction. For example, if the strike runs north-south and the dip is 30˚, it would be necessary to specify "to the west" or "to the east." Similarly if the strike is northeast-southwest and the dip is 60˚, it would be necessary to say "to the northwest" or "to the southeast." In the case of the vertical wall with a dip angle of 90˚, there is no dip direction. The dip points straight down, not toward any compass direction.

Measurement of geological features is done with a special compass that has a built-in **clinometer**, which is a device for measuring vertical angles. The strike is measured by aligning the compass along a horizontal line on the surface of the feature (Figure 13.36, left). The dip is measured by turning the compass on its side and aligning it along the dip direction (Figure 13.36, right).

Figure 13.36 Measurement of strike (left) and dip (right) using a geological compass with a clinometer. *Source: Steven Earle (2015) CC BY 4.0*

Strike and dip are used to describe any other planar features, including joints, faults, dykes, sills, and even the foliation planes in metamorphic rocks. Figure 13.37 shows an example of how we would depict the beds that make up an anticline on a map. The beds on the west (left) side of the map are dipping at various angles to the west. The beds on the east side are dipping to the east. The beds in the middle are horizontal; this is denoted by a cross within a circle on the map. The dyke is dipping at 80˚ to the west. The hinge line of the fold is denoted with a dashed line on the map, with two arrows pointing away from it, indicating the general dip directions of the limbs. If it were a syncline, the arrows would point inward toward the line.

Figure 13.37 A depiction of an anticline and a dyke in cross-section (looking from the side) and in map view (or plan view) with the appropriate strike-dip and anticline symbols. *Source: Steven Earle (2015) CC BY 4.0*

Exercise: Putting Strike & Dip on a Map

This cross-section shows seven tilted sedimentary layers (a to g), a fault, and a steeply dipping dyke.

Figure 13.38 Practice with strike and dip symbols. *Source: Steven Earle (2015) CC BY 4.0*

1. Place strike and dip symbols on the map to indicate the orientations of the beds shown, the fault, and the dyke.
2. What type of fault is shown?
3. What kind of stress created the fault?

Chapter 13 Summary

The topics covered in this chapter can be summarized as follows:

13.1 Stress and Strain

Stress within rocks, which includes compression, extension and shearing, originates from plate tectonic processes and the weight of overlying rocks. Rock that is stressed responds with either elastic or plastic strain, and may eventually break. The way a rock responds to stress depends on its composition and structure, the rate at which strain is applied, and also on the temperature, pressure, and the presence of fluid within the rock.

13.2 Folding

Folding is generally a ductile response to compression, although some brittle behaviour can happen during folding. A fold with a hinge that points upward is an anticline. A fold with a hinge that points downward is a syncline. The axial surface of a fold can be vertical, inclined, or even horizontal. The landforms produced by folds will depend on the resistance to weathering of rock layers within the fold.

13.3 Fractures, Joints, and Faults

Joints typically form during extension, but can also form during compression. Faulting, which involves the displacement of rock, can take place during compression or extension, as well as during shearing at transform boundaries. Thrust faults are a type of reverse fault with a fault plane tilted at a low angle. Thrust faults are common in mountain belts formed by plate collisions.

13.4 Mountain Building

Mountain building in zones of plate collision is called orogeny. The mountains that form are orogens, and consist of crust thickened and deformed folding and faulting, as well as the intrusion of igneous rocks. Orogens in ocean-continent collision zones have volcanoes. Mountains formed in rift zones are the result of tilting of normal-faulted blocks, or some normal-faulted blocks subsiding while others remain elevated.

13.5 Measuring Geological Structures

The strike and dip of planar surfaces, such as a bedding planes, fractures or faults are measured to help understand the geological history of a region. Special symbols are used to show the orientation of structural features on geological maps.

Review Questions

1. 1. What types of plate boundaries are most likely to contribute to (a) compression, (b) extension, and (c) shearing?
2. 2. Explain the difference between elastic strain and plastic strain.
3. 3. List some of the factors that influence whether a rock will undergo ductile deformation or break when placed under stress.
4. 4. Draw in the axial traces of the folds in Figure 13.39, and label each with the appropriate type (e.g., overturned syncline).

Figure 13.39 A cross-section showing folds. *Source: Steven Earle (2015) CC BY 4.0*

5. Explain why fractures are common in volcanic rocks.

6. What is the difference between a normal fault and a reverse fault, and under what circumstances would you expect these to form?

7. What type of fault would you expect to see at a transform plate boundary?

8. Figure 13.40 is a map of the geology of a region. The coloured areas represent sedimentary beds.

(i) Describe in words the *general* attitude (strike and dip) of these beds.

(ii) What is “a” and what is its attitude?

(iii) What is “b” and what is its attitude?

(iv) Is "b" left handed or right handed?

Figure 13.40
Geological map.
Source: Steven Earle (2015) CC BY 4.0

Answers at the end of the chapter

References

Heard, H. C. (1960). Transition from Brittle Fracture to Ductile Flow in Solenhofen Limestone as a Function of Temperature, Confining Pressure, and Interstitial Fluid Pressure. In D. Griggs & D. Handin (Eds.), *Rock Deformation (A Symposium): Geological Society of America Memoir 79* (pp. 193-226). https://doi.org/10.1130/MEM79

Symonds, W. S. (1872). *Records of the rocks; or, Notes on the geology, natural history, and antiquities of North & South Wales, Devon, & Cornwall.* London: J.Murray https://archive.org/details/recordsofrocksor00symoiala

Answers to Review Questions

1. Convergent plate boundaries are the most likely to contribute to compression, divergent boundaries to extension, and transform boundaries to shearing. However, all of these stress regimes can exist at any one of these boundaries.

2. When elastic strain takes place the rock can rebound to its original shape when stress is removed. Plastic strain is permanent.

3. Ductile deformation is more likely under higher temperatures and confining pressures. It is more likely when rocks are deformed slowly, and by compression. It is more likely for sedimentary rocks, and for rocks without fluids.

4. Axial traces are shown with dashed red lines.

Figure 13.41 Folds with labels. *Source: Steven Earle (2015) CC BY 4.0*

5. Volcanic rocks cool quickly at surface and the resulting reduction in volume can easily lead to fracturing.

6. In a normal fault the rock above the fault (hanging wall or headwall) moves down with respect to the lower rock (footwall). This normally indicates extension. In a reverse fault the hanging wall is pushed up, which indicates compression.

7. Most faults near transform boundaries are strike-slip faults, meaning that there is horizontal motion along the fault.

8.(i) The beds are dipping at about 30˚ to the northwest. (ii) "a" is a dyke and it is dipping steeply to the northeast. (iii) "b" is a fault and it is dipping steeply to the southeast. (iv) The motion on fault "b" appears to be left handed.

Figure 13.42 Geological map with strike and dip symbols. *Source: Steven Earle (2015) CC BY 4.0*

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-SA-NC 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 14. Streams and Floods

Adapted by Joyce M. McBeth, University of Saskatchewan
from *Physical Geology* by Steven Earle

Learning Objectives

After carefully reading this chapter, completing the exercises within it, and answering the questions at the end, you should be able to:

- Explain the hydrological cycle, its relevance to streams, and describe the residence time of water in these systems
- Describe what a drainage basin is, and explain the origins of the different types of drainage patterns
- Explain how streams become graded, and how certain geological and anthropogenic changes can result in a stream becoming ungraded
- Describe the formation of stream terraces
- Describe the processes that move sediments in streams, and how changes in stream velocity affect the types of sediments that are moved by the stream
- Explain the origin of natural stream levees
- Describe the process of stream evolution and the types of environments where one would expect to find straight-channel, braided, and meandering streams
- Describe the annual flow characteristics of typical streams in Canada and the processes that lead to flooding
- Describe some of the important historical floods in Canada
- Determine the probability of floods of various magnitudes, based on the flood history of a stream
- Explain some of the steps that we can take to limit damage from flooding

Why Study Streams?

Figure 14.1 A small waterfall on Johnston Creek in Johnston Canyon, Banff National Park, AB
Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

Streams are the most important agents of erosion and transportation of sediments on Earth's surface at this time in Earth's history. They are responsible for generating much of the topography on that land surfaces that we see around us. They are places of beauty and tranquility, and they provide much of the water that is essential to our existence. But streams are not always peaceful and soothing. During large storms and rapid snowmelts, they can become raging torrents capable of moving cars and houses, and destroying roads and bridges. When they spill over their banks, they can flood huge areas, devastating populations and infrastructure. Over the past century, many of the most damaging natural disasters in Canada have been floods.

15.1 The Hydrological Cycle

Water is constantly on the move. It is evaporated from the oceans, lakes, streams, the surface of the land, and from plants (transpiration) by solar energy (Figure 14.2). It is transported in its gaseous form through the atmosphere by the wind and condenses to form clouds of water droplets or ice crystals. It falls to the Earth's surface as rain or snow and flows through streams, into lakes, and eventually back to the oceans. Water on the surface and in streams and lakes infiltrates the ground to become groundwater. Groundwater slowly moves through soils, surficial materials, and pores and cracks in the rock. The groundwater flow paths can intersect with the surface and the water can then move back into streams, lakes and oceans.

Figure 14.2 The various components of the water cycle. Black or white text indicates the movement or transfer of water from one reservoir to another. Yellow text indicates the storage of water. Source: Steven Earle (2015) CC BY-SA 3.0 view source https://opentextbc.ca/geology/ after Wikimedia user "Ingwik" (2010) CC-BY-SA 3.0. view source http://bit.ly/HydCyc

Water is stored in various reservoirs as it moves across and through the Earth. A **reservoir** is a space that stores water. It can be a space we can easily visualize (such as a lake) or a space that is more difficult to visualize (such as the atmosphere or the groundwater in a region). The largest reservoir is the ocean, accounting for 97% of the total volume of water on Earth (Figure 14.3). Ocean water is salty, but the remaining 3% of water on Earth is fresh water. Two-thirds of our fresh water is stored in the ground and one-third is stored in ice. The remaining fresh water (about 0.03% of the total) is stored in lakes, streams, vegetation, and the atmosphere.

To put these percentages in perspective, we can compare a 1 litre container of water to the entirety of the Earth's water supply (Figure 14.4). We start by almost filling the container with 970 ml of water and 34 g of salt, to simulate all the sea water on Earth. Then we add one regular-sized (*ca* 20 ml) ice cube (representing glacial ice) and two teaspoons (*ca* 10 ml) of groundwater. All of the water that we see around us in lakes and streams and in the atmosphere can be represented by adding three more drops of water from an eyedropper.

Although the water in the atmosphere is only a small proportion of the total water on Earth, the volume is still very large. At any given time, there is the equivalent of approximately 13,000 km^3 of water in the air in the form of water vapour and water droplets in clouds. Water is evaporated from the oceans, vegetation, and lakes at a rate of 1,580 km^3 per day, and each day nearly the same volume falls back as rain and snow over the oceans and land. Most of the precipitation that falls onto land returns to the ocean in the form of stream flow (117 km^3/day) and groundwater flow (6 km^3/day). Most of the rest of this chapter is about this 117 km^3/day of streamflow.

Figure 14.3 The storage reservoirs for water on Earth. Glacial ice is represented by the white band, groundwater the red band, and surface water the very thin blue band at the top. The 0.001% stored in the atmosphere is not shown. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/, data from USGS Water Science School (2016) http://bit.ly/USGSH2O

Figure 14.4 Representation of the Earth's water in a 1 litre container. The three drops represent all of the fresh water in lakes, streams, and wetlands, plus all of the water in the atmosphere. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

Exercise 14.1 How Long Does Water Stay in the Atmosphere?

The residence time of a water molecule in the atmosphere (or any of the other reservoirs) can be estimated by dividing the total amount of water in the reservoir by the rate at which it is removed. For the atmosphere, we know that the reservoir size is 13,000 km^3, and the rate is 1,580 km^3/day. If we divide 13,000 km^3 by 1,580 km^3/day, we get 8.22 days. This means that, *on average*, a molecule of water stays in the atmosphere for just over eight days. "Average" needs to be emphasized here because some molecules remain in the air for only a few hours, while others may remain in the air for weeks.

The volume of the oceans is 1,338,000,000 km^3 and the rate of removal of water from the oceans is approximately the same as the atmosphere (1,580 km^3/day). What is the average residence time of a water molecule in the ocean?

14.2 Drainage Basins

A **stream** is a body of flowing surface water of any size, ranging from a tiny trickle to a mighty river. The area from which the water flows to form a stream is known as its **drainage basin** or **watershed**. All of the precipitation (rain or snow) that falls within a drainage basin eventually flows into its stream, unless some of this water is able to cross into an adjacent drainage basin via groundwater flow. An example of a drainage basin is shown in Figure 14.5.

Figure 14.5 A schematic diagram of the drainage basin of Cawston Creek near Keremeos, BC. The blue line shows the extent of the drainage basin. The dashed red line is the drainage basin of one of its tributaries. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

An important characteristic of streams is their **gradient**, the rate of change in elevation with distance along the stream. A steep gradient has a rapid change in elevation with horizontal distance, and a shallow gradient has a slow change in elevation with horizontal distance. Cawston Creek drainage basin in BC is approximately 25 km^2 and is a typical small drainage basin within a very steep glaciated valley. As shown in Figure 14.6, the upper and middle parts of the creek have steep gradients averaging about 200 m/km but ranging from 100 to 350 m/km, and the lower part, within the valley of the Similkameen River, is relatively flat at <5 m/km.

The shape of the valley has changed through time to result in the shape we see now. First, there was tectonic uplift (related to tectonic plate convergence). Then the landscape changed due to stream erosion and mass wasting (landslides). This was followed by several episodes of glacial erosion. Finally, there was post-glacial stream erosion up to the present time. The lowest elevation of Cawston Creek (275 m, where the creek flows into the Similkameen River) is its **base level**. Cawston Creek cannot erode below this level unless the Similkameen River erodes deeper into its flood plain (the area that is inundated during a flood). Base level is the elevation where a stream will no longer erode deeper into the bedrock or sediments it flows through, because the

elevation of the stream does not drop below this level, and further erosion can only occur if there is an elevation drop to propel the water deeper due to the force of gravity.

The ocean is the ultimate base level, but lakes and other rivers act as base levels for many smaller streams.

Figure 14.6 Profile of the main portion of Cawston Creek near Keremeos, BC. The maximum elevation of the drainage basin is about 1,840 m, near Mount Kobau. The base level is 275 m, at the Similkameen River. As shown, the gradient of the stream can be determined by dividing the change in elevation between any two points (rise) by the distance between those two points (run). Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

The water supply for the Vancouver, BC metropolitan area comes from three large drainage basins on the north shore of Burrard Inlet, as shown in Figure 14.7. This map illustrates the concept of a drainage basin divide. The boundary between two drainage basins is the ridge of land between them. A drop of rain falling on the boundary between the Capilano and Seymour drainage basins, for example, could flow into either basin. Rain falling on the Capilano basin side cannot flow into the Seymour drainage basin, because of the drainage basin divide.

Figure 14.7 The three drainage basins that supply water to the metropolitan Vancouver, BC area. Source: Wikimedia user "Alaidlaw" (2016) CC BY-SA 2.0. view source https://en.wikipedia.org/wiki/Metro_Vancouver_watersheds#/media/File:Metro_Vancouver_Watershed_Boundaries.jpg

The pattern of tributaries within a drainage basin depends largely upon the type of underlying rock, and on structures within that rock such as folds, fractures, and faults. Three types of drainage patterns are illustrated in Figure 14.8. **Dendritic** patterns, which are by far the most common, develop in areas where the rock (or unconsolidated material) beneath the stream does not have structures that control the stream flow patterns such as folds and joints; the materials can be eroded by the stream equally easily in all directions. Most areas of British Columbia have dendritic patterns, as do most areas of the prairies and the Canadian Shield.

Trellis drainage patterns typically develop where sedimentary rocks have been folded or tilted, and then eroded to varying degrees depending on their resistance to erosion. The Rocky Mountains of BC and AB have some fine examples of trellis drainage.

Rectangular patterns develop in areas that have very little topography and a system of bedding planes, fractures, or faults that form a rectangular network. Rectangular drainage patterns are rare in Canada.

Figure 14.8 Typical dendritic, trellis, and rectangular stream drainage patterns. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

In many parts of Canada, especially relatively flat areas with thick glacial sediments, and throughout much of Canadian Shield in eastern and central Canada, drainage patterns are chaotic, also known as **deranged** (Figure 14.9, left). Lakes and wetlands are common in this type of environment.

Radial drainage (Figure 14.9, right) is a pattern that forms around isolated mountains (such as volcanoes) or hills. The individual streams that radiate out from the hill typically have dendritic drainage patterns.

Figure 14.9 Left: a typical deranged pattern; right: a typical radial drainage pattern developed around a mountain or hill. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

The process of a stream eroding downward into bedrock is called **downcutting**. Over geological time, and during tectonic quiescence, a stream will erode its drainage basin into a smooth profile similar to that shown in Figure 14.10. This is called a **graded stream**. Graded streams are steepest in their headwaters and their gradient gradually decreases toward their mouths. **Ungraded streams** are still in the

process of rapidly eroding and downcutting their drainage basin, they have steep sections at various points, and typically have rapids and waterfalls at numerous locations along their lengths (e.g., Cawston Creek, Figure 14.6).

Figure 14.10 The topographic profile of a typical graded stream. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

A graded stream can become ungraded if there is renewed tectonic uplift, or if there is a change in the base level. Base level changes can occur due to tectonic uplift or some other reason such as construction of a dam downstream. As stated earlier, the base level of Cawston Creek is defined by the level of the Similkameen River, but this can change, and has done so in the past. Figure 14.11 shows the valley of the Similkameen River in the Keremeos area. The river channel is just beyond the row of trees. The green field in the distance is underlain by material eroded from the hills behind and deposited by a small creek (not Cawston Creek) adjacent to the Similkameen River when its level was higher than it is now. Some time in the past several centuries, the Similkameen River eroded down through these deposits (forming the steep bank on the other side of the river), and the base level of the small creek was lowered by about 10 m. Over the next few centuries, this creek will erode through the sediments and eventually become graded again.

Figure 14.11 An example of a change in the base level of a small stream that flows into the Similkameen River near Keremeos. The previous base level was near the top of the sandy bank. The current base level is the Similkameen River, located on the far side of the line of trees. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

Another example of a change in base level can be seen along the Juan de Fuca Trail on southwestern Vancouver Island. As shown in Figure 14.12, many of the small streams along this part of the coast flow into the ocean as waterfalls. It is evident that the land in this area has risen by about 5 m in the past few thousand years, probably in response to deglaciation. The streams that used to flow directly into the ocean now have a lot of downcutting to do before they will be a graded stream again.

Figure 14.12 Two streams with a lowered base level on the Juan de Fuca Trail, southwestern Vancouver Island. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

Sediments accumulate within the channel and flood plain of a stream, and then, if the base level changes, or if there is less sediment supplied to the stream to deposit, the stream may cut down through these existing sediments to form terraces. A terrace on the Similkameen River is shown in Figure 14.11 and terraces on the Fraser River are shown in Figure 14.13.

Figure 14.13 Terraces on the Fraser River north of Lillooet, BC (above the river on the left-hand side of the image). Source: Wikimedia user "China Crisis" (2007) CC BY-SA 3.0. view source https://en.wikipedia.org/wiki/File:FraserRiverNearLillooet.jpg

In the late 19th century, American geologist William Davis proposed that streams and the surrounding terrain develop in a cycle of erosion (Figure 14.14). Following tectonic uplift, the stream patterns are immature. Streams erode quickly, developing deep V-shaped valleys that tend to follow relatively straight paths. Gradients are high, and profiles are ungraded. Rapids and waterfalls are common. As the landscape matures, the streams erode wider valleys and deposited thick sediment layers. Even after maturity, gradients are slowly reduced and grading increases. In old age, streams are surrounded by rolling hills, and they occupy wide sediment-filled valleys. Meandering patterns are common, and erosion now is focussed towards the channel walls, with little downcutting.

Figure 14.14 A depiction of the Davis cycle of erosion: a: initial stage, b: youthful stage, c: mature stage, and d: old age. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

Davis's work was done long before the idea of plate tectonics, and he was also not familiar with the impacts of glacial erosion on streams and their environments. While some parts of his idea are out of date, it is still a useful way to understand streams and their evolution. Plate tectonic activity and other processes such as isostatic rebound after glaciation results in uplift that alters stream gradients, so streams are constantly adjusting due to these changing conditions. It would be relatively rare to find a stream that is able to mature through all of these stages without interruption.

Exercise 14.2 The Effect of a Dam on Base Level

When a dam is built on a stream, a reservoir (artificial lake) forms behind the dam. The dam reservoir temporarily (for many decades at least) creates a new base level for the part of the stream above the reservoir. How does the formation of a dam reservoir affect the stream where it enters the reservoir, and what happens to the sediment it was carrying? The water leaving the dam has no sediment in it. Why? How does this affect the stream below the dam?

Revelstoke Dam and Revelstoke Lake on the Columbia River at Revelstoke, BC. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

14.3 Stream Erosion and Deposition

Stream Velocity Depends on the Shape and Size of the Channel

Flowing water is a very important mechanism for both erosion and deposition. Water flow in a stream is primarily related to the stream's gradient, but it is also controlled by the geometry of the stream channel. As shown in Figure 14.15, water flow velocity decreases due to friction along the stream bed. The stream is thus slowest at the bottom and edges and fastest near the surface and in the middle of the stream (where there is the least amount of friction). The velocity just below the surface of the water is typically a little higher than right at the surface because of friction between the water and the air. On a curved section of a stream, flow is fastest on the outside of the curve and slowest on the inside of the curve.

Figure 14.15 The relative velocity of stream flow depending on whether the stream channel is straight or curved (left). (Right) it is also dependent on the water depth. The length of each of the arrows indicates the relative velocity of the stream at that position in the channel. Shorter arrows mean slower flow. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

Another important factor influencing stream-water velocity is the **discharge**, or volume of water passing a point in a unit of time (e.g., m^3/second). Water levels rise during a flood and due to the higher discharge of water the stream flow velocity increases. The higher discharge also increases the cross-sectional area of the stream, so it fills up the channel. In a flood it may overflow the banks. Another factor that affects stream-water velocity is the size of sediments on the stream bed. Large particles tend to slow the flow more than small ones.

Sediment Transport Depends on Stream Velocity and Turbulence

If you drop a piece of gravel into a glass of water, it will sink quickly to the bottom. If you drop a grain of sand into the same glass, it will sink more slowly. A grain of silt will take longer yet to get to the bottom, and a particle of fine clay will take a long time settle out. The rate of settling is determined by the balance between gravity and friction, as shown in Figure 14.16.

One of the key principles of sedimentary geology is that the ability of a moving medium (air or water) to move sedimentary particles and keep them moving is dependent on the velocity of flow. The faster the medium flows, the larger the particles it can move. As you probably know from intuition and from experience, streams that flow rapidly tend to be **turbulent** (flow paths are chaotic and the water surface appears rough) and the water

may be muddy. In contrast, streams that flow more slowly tend to have **laminar flow** (straight-line flow and a smooth water surface) and clearer water. Turbulent flow is more effective than laminar flow at keeping sediments suspended within the water.

Particles within a stream are transported in different ways depending on their size (Figure 14.17). Large particles which rest on the stream bed are known as the **bedload.** The bedload may only be transported when the flow rate is rapid and under flood conditions. They are transported by **saltation** (bouncing along, and colliding with other particles) and by **traction** (being pushed along by the force of the flow).

Smaller particles may rest on the bottom occasionally, where they can be transported by saltation and traction, but they can also be held in suspension in the flowing water (the **suspended load**), especially at higher flow velocities.

Stream water also has a **dissolved load**, which represents (on average) about 15% of the mass of material transported and includes ions such as calcium (Ca^{+2}) and chloride (Cl-) in solution. The solubility of these ions is not affected by flow velocity.

Figure 14.16 How quickly a grain settles to the bottom of a stream depends on its mass (affecting the force of gravity acting on it), and the friction between the grain and the water, which slows the fall of the grain. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

Figure 14.17 Modes of transportation of sediments and dissolved ions (represented by red dots with + and – signs) in a stream. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

If you look at a typical stream, there are always some sediments being deposited, some staying where they are, and some being eroded and transported. This changes over time as the discharge of the river changes in response to changing weather conditions.

The Hjulström-Sundborg Diagram Summarizes What Happens to Grains of Different Sizes at Different Stream Velocities

The faster water is flowing, the larger the particles that can be kept in suspension and transported within the flowing water. However, as Swedish geographer Filip Hjulström discovered in the 1940s, the relationship between grain size and the likelihood of a grain being eroded, transported, or deposited is not as simple as one might imagine (Figure 14.18). Consider, for example, a 1 mm grain of sand. If it is resting on the bottom of the stream, it will remain there until the flow velocity is high enough to erode it (*ca* 20 cm/s). But once it is in suspension, that same 1 mm particle will remain in suspension as long as the velocity doesn't drop below 10 cm/s. For a 10 mm gravel grain, the velocity is 105 cm/s to be eroded from the bed but only 80 cm/s to remain in suspension.

Figure 14.18 The Hjulström-Sundborg diagram showing the relationships between particle size and the tendency to be eroded, transported, or deposited, at different current velocities Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

On the other hand, a 0.01 mm silt particle only needs a velocity of 0.1 cm/s to remain in suspension, but requires 60 cm/s to be eroded. In other words, a tiny silt grain requires a greater velocity to be eroded than a grain of sand that is 100 times larger! For clay-sized particles, the discrepancy is even greater. In a stream, the most easily eroded particles are small sand grains between 0.2 mm and 0.5 mm. Anything smaller or larger requires a higher water velocity to be eroded and entrained in the flow. The reason for this is that small particles, especially tiny grains of clay, possess a net surface charge, hence experience a strong tendency to stick together, and so are difficult to erode from the stream bed.

It is important to be aware that a stream can both erode and deposit sediments at the same time. At 100 cm/s, for example, silt, sand, and medium gravel will be eroded from the stream bed and transported in suspension, coarse gravel will be transported by saltation and traction, pebbles will be both transported by saltation and traction, and will also be deposited, and cobbles and boulders will remain stationary on the stream bed.

Exercise 14.3 Understanding the Hjulström-Sundborg Diagram

Refer to the Hjulström-Sundborg diagram in Figure 14.18 to answer these questions.

1. A fine sand grain (0.1 mm) is resting on the bottom of a stream bed.
   a) What stream velocity will it take to get this sand grain into suspension?
   b) Once the particle is in suspension, the velocity starts to drop. At what velocity will it finally come back to rest on the stream bed?
2. A stream is flowing at 10 cm/s (which means it takes 10 s to travel 1 m).
   a. What size of particles can be eroded at 10 cm/s?
   b. What is the largest particle that, once in suspension, will remain in suspension at 10 cm/s?

Natural Levees Form Because of Changes in Stream Velocity

A stream typically reaches its greatest velocity when it is close to flooding over its banks. This is known as the **bank-full stage**, as shown in Figure 14.19. When the flooding stream overtops its banks and occupies the wide area of its flood plain, the water has a much larger area to flow through and the velocity drops dramatically. As water flows from the channel out across the flood plain, it slows down and starts to deposit its sediment load. This forms an elevated bank known as a **levee** along the edges of the channel**.** The coarsest and thickest sediments are deposited near the channel banks, with particle size and thickness decreasing as you move further into the flood plain. People also build levees as flood control measures; the idea for this engineered solution to floods came from the naturally-build levees that form during floods.

Figure 14.19 The development of natural levees during flooding of a stream. The sediments of the levee become increasingly fine away from the stream channel, and even finer sediments — clay, silt, and very fine sand — are deposited across most of the flood plain. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

14.4 Stream Types

Stream channels can be straight or curved, deep or shallow, cleared or filled with coarse sediments. The cycle of erosion has some influence on the nature of a stream, but there are several other factors that are important.

Figure 14.20 The Cascade Falls area of the Kettle River, near Christina Lake, BC. This stream has a step-pool morphology and a deep bedrock channel. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

Youthful streams that are actively downcutting their channels tend to be relatively straight and are typically ungraded (meaning that rapids and waterfalls are common). As shown in Figures 14.1 and 14.20, youthful streams commonly have a **step-pool** morphology, meaning that the stream consists of a series of pools connected by rapids and waterfalls. They also have steep gradients, and steep and narrow V-shaped valleys. In some cases these valley walls are steep enough to be called canyons.

In mountainous terrain, such as that in western AB and BC, steep youthful streams typically flow into wide and relatively low-gradient U-shaped glaciated valleys. The youthful streams have high sediment loads, and when they flow into the lower-gradient glacial valleys where the velocity is no longer high enough to carry all of the sediment, **braided** stream patterns develop, characterized by a series of narrow channels separated by gravel bars (Figure 14.21).

Figure 14.21 The braided channel of the Kicking Horse River at Field, BC. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

Braided streams can develop anywhere where there is more sediment than a stream is able to transport. One such environment is in volcanic regions, where explosive eruptions produce large amounts of unconsolidated material that gets washed into streams. The Coldwater River next to Mt. St. Helens in Washington State is a good example of such a braided stream (Figure 14.22).

A stream that occupies a wide, flat flood plain with a low gradient typically carries only sand-sized and finer sediments and develops a sinuous flow pattern. As you saw in Figure 14.15, when a stream flows around a bend, the main current of the stream flows near the outside portion of the bend. This leads to erosion of the banks on the outside of the bend, and deposition of a **point bar** on the inside of the bend (Figure 14.23). Over time, the sinuosity of the stream becomes increasingly exaggerated, and the channel migrates throughout its flood plain, forming a **meandering** pattern.

Figure 14.22 The braided Coldwater River, Mt. St. Helens, Washington. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

Figure 14.23 The meandering channel of the Bonnell Creek, Nanoose, BC. The stream is flowing toward the viewer. The sand and gravel point bar must have formed when the creek was higher and the flow faster than it was when the photo was taken, as the current stream velocity is too low to carry such coarse sediments. Most erosion and deposition take place during flooding events. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

A well-developed meandering river is shown in Figure 14.24. The meander in the middle of the photo has reached the point where the thin neck of land between two parts of the channel is about to be eroded through. When this happens, an **oxbow** lake will form. These are small cut off bends from earlier curves in the river; several are visible outside the path of the main stream in Figure 14.24.

At the point where a stream enters a body of water such as a lake or the ocean, the flow rates drops dramatically, and sediment is deposited. Over time, as more and more sediments are deposited, the sediments form a distinctive triangular shape (with the bottom broad part of the triangle facing the ocean or lake and the point of the triangle facing upstream). This is called a **delta**; these are named after the Greek letter delta which is in the shape of a triangle. The Fraser River has created a large delta in BC where the river meets the Strait of Georgia (Figure 14.25). Much of the Fraser delta is very young in geological terms. Shortly after the end of the last glaciation (10,000 years ago), the delta did not extend past New Westminster. Since that time, all of the land that makes up Richmond, Delta, and parts of New Westminster and south Surrey has formed from sediment depositing from the Fraser River. You can see a more detailed description of the Fraser delta on the Geoscape Vancouver website: http://www.cgenarchive.org/vancouver-fraserdelta.html

Figure 14.24 The meandering channel of the Nowitna River, Alaska. Numerous oxbow lakes are present, and another meander cutoff will soon take place. Source: Oliver Kumis (2002) CC-BY-SA 2.0 view source http://bit.ly/1SmQL7B

Figure 14.25 The delta of the Fraser River and the plume of sediment that extends across the Strait of Georgia. The land outlined in red has formed over the past 10,000 years. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/ after NASA, September 2011, http://bit.ly/FrasR

Exercise 14.4 Calculating Stream Gradients

The gradient is the key factor controlling stream velocity, and stream velocity controls sediment erosion and deposition. This map shows the elevations of Priest Creek in the Kelowna area. The length of the creek between 1,600 m and 1,300 m elevation is 2.4 km, so the gradient is (1,600 m - 1,300 m)/2.4 km = 125 m/km.

Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

1. Use the scale bar to estimate the distance between 1,300 m and 600 m and calculate the gradient between these two elevations.
2. Estimate the gradient between 600 and 400 m.
3. Estimate the gradient between 400 m on Priest Creek and the point where Mission Creek enters Okanagan Lake.

14.5 Flooding

The discharge levels of streams are highly variable depending on the time of year and on variations in the weather from one year to the next. In Canada, most streams show discharge variability similar to that of the Stikine River in northwestern BC, illustrated in Figure 14.26. The Stikine River has its lowest discharge levels in the depths of winter when freezing conditions persist throughout most of its drainage basin. Discharge starts

to rise slowly in May, and then rises dramatically through the late spring and early summer as the winter snow melts. For the year shown, the minimum discharge of the Stikine River was 56 m^3/s in March, and the maximum was 37 times higher at 2,470 m^3/s in May.

Figure 14.26 Variations in discharge of the Stikine River during 2013. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/ from data at Water Survey of Canada, Environment Canada, http://www.ec.gc.ca/rhc-wsc/

Streams in coastal areas of southern British Columbia show a very different pattern from those in most of the rest of the country. In this region, the drainage basins receive a lot of rain (rather than snow) during the winter and also do not remain entirely frozen throughout the winter. The Qualicum River on Vancouver Island typically has its highest discharge levels in January or February and its lowest levels in late summer (Figure 14.27). In 2013, the minimum discharge of the Qualicum River was 1.6 m^3/s in August, and the maximum was 34 times higher at 53 m^3/s in March.

Figure 14.27 Variations in discharge of the Qualicum River during 2013. Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/ from data at Water Survey of Canada, Environment Canada, http://www.ec.gc.ca/rhc-wsc/

When a stream's discharge increases, both the water level (**stage**) and the velocity increase as well. Rapidly flowing streams become muddy, and large volumes of sediment are transported both in suspension and along the stream bed. In extreme situations, the water level reaches the top of the stream's banks (the **bank-full stage**, see Figure 14.19), and if it rises further, it will overflow the banks and floods the surrounding terrain. In the case of mature or old-age streams, this could include a vast area of relatively flat ground known as a flood plain, which

is the area that is typically covered with water during a major flood. Since fine river sediments are deposited on flood plains, they are ideally suited for agriculture, and thus are typically occupied by farms and residences, and in many cases, by towns or cities. Such infrastructure is highly vulnerable to damage from flooding, and the people that live and work there are at risk.

Most streams in Canada have the greatest risk of flooding in the late spring and early summer when stream discharges rise in response to melting snow. In some cases, this is exacerbated by spring storms. In years when melting is especially fast and/or spring storms are particularly intense, flooding can be very severe.

One of the worst floods in Canadian history took place in the Fraser Valley of BC in late May and early June of 1948. The early spring of that year had been cold, and a large snow pack in the interior was slow to melt. In mid-May, temperatures rose quickly and melting was accelerated by rainfall. Fraser River discharge levels rose rapidly over several days during late May, and the dykes built to protect the valley were breached in a dozen places. Approximately one-third of the flood plain was inundated, and many homes and other buildings were destroyed, but there were no deaths.

The Fraser River flood of 1948, which was the worst flood in the Fraser Valley in the past century, was followed by very high river levels in 1950 and 1972, and by relatively high levels several times since then, the most recent being 2007 (Table 13.1). In the years following 1948, millions of dollars were spent repairing and raising the existing dykes and building new ones. Since then damage from flooding in the Fraser Valley has been relatively limited.

Rank	Year	Month	Date	Stage (m)	Discharge (m3/s)
1	1948	May	31	11.0	15,200
2	1972	Jun	16	10.1	12,900
3	1950	Jun	20	9.9	12,500
4	1964	Jun	21	9.6	11,600
5	1997	Jun	5	9.5	11,300
6	1955	Jun	29	9.4	11,300
7	1999	Jun	22	9.4	11,000
8	2007	Jun	10	9.3	10,850
9	1974	Jun	22	9.3	10,800
10	2002	Jun	21	9.2	10,600

Table 14.1 Ranking of the maximum stage and discharge values for the Fraser River at Hope between 1948 and 2008. Typical discharge levels are ~1,000 m^3/s. Source: Data from Mannerstrom (2008) Comprehensive Review of Fraser River at Hope Flood Hydrology and Flows Scoping Study, Report prepared for the B.C. Ministry of the Environment. Available at: http://www.env.gov.bc.ca/wsd/public_safety/flood/pdfs_word/review_fraser_flood_flows_hope.pdf

Serious flooding occurred in July 1996 in the Saguenay-Lac St. Jean region of Quebec. In this case, the floods were caused by two weeks of heavy rainfall followed by one day of exceptional rainfall. On July 19, 1996 there was 270 mm of rain, equivalent to the region's normal rainfall for the entire month of July. Ten deaths were attributed to the Saguenay floods, and the economic toll was estimated at $1.5 billion.

Just a year after the Saguenay floods, the Red River in Minnesota, North Dakota, and Manitoba reached its highest level since 1826. As is typical for the Red River, the 1997 flooding was due to rapid snowmelt. Due to the south to north flow of the river, the flooding starts in Minnesota and North Dakota, where melting begins

earlier, then extends northwards. The residents of Manitoba had plenty of warning that the 1997 flood was coming because there was severe flooding at several locations on the U.S. side of the border.

After the 1950 Red River flood, the Manitoba government built a channel around the city of Winnipeg to reduce the potential of flooding in the city (Figure 14.28). Known as the Red River Floodway, the channel was completed in 1964 at a cost of $63 million. Since then it has been used many times to alleviate flooding in Winnipeg, and is estimated to have saved many billions of dollars in flood damage. The massive 1997 flood (Figure 14.28, right side) was almost too much for the floodway; in fact the amount of water diverted was greater than the designed capacity. The floodway has recently been expanded so that it can be used to divert even more of the Red River's flow away from Winnipeg.

Figure 14.28 (left) map of the Red River Floodway around Winnipeg, Manitoba; (right) aerial view of the southern (inlet) end of the floodway during the 1997 Red River flood. Sources: (left) Wikimedia user "Kmusser" (2007) CC BY 2.5 view source http://en.wikipedia.org/wiki/1997_Red_River_Flood#/media/File:Rednorthfloodwaymap.png. (right) Natural Resources Canada 2012, courtesy of the Geological Survey of Canada (Photo 2000-118 by G.R. Brooks).

Canada's most costly flood ever was the June 2013 flood in southern Alberta. The flooding was initiated by snowmelt and worsened by heavy rains in the Rockies due to an anomalous flow of moist air from the Pacific and the Caribbean. At Canmore, AB rainfall amounts exceeded 200 mm in 36 hours, and at High River, AB 325 mm of rain fell in 48 hours.

In late June and early July, the discharges of several rivers in the area, including the Bow River in Banff, Canmore, and Exshaw, the Bow and Elbow Rivers in Calgary, the Sheep River in Okotoks, and the Highwood River in High River, reached levels that were 5 to 10 times higher than normal for that time of year (see Exercise 14.5). Large areas of Calgary, Okotoks, and High River were flooded, and five people died (see Figures 14.29 and 14.30). The cost of the 2013 flood is estimated to be approximately $5 billion.

Figure 14.29 Map of the communities most affected by the 2013 Alberta floods (in orange) Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/

Figure 14.30 Flooding in Calgary (June 21, left) and Okotoks (June 20, right) during the 2013 southern Alberta flood. Sources: (left) Wikimedia user Ryan L.C. Quan (2013) CC BY-SA 3.0 view source https://en.wikipedia.org/wiki/2013_Alberta_floods#/media/File:Riverfront_Ave_Calgary_Flood_2013.jpg (right) Wikimedia user "Stephanie N. Jones" (2013) CC BY-SA 3.0 view source https://en.wikipedia.org/wiki/2013_Alberta_floods#/media/File:Okotoks_-_June_20,_2013_-_Flood_waters_in_local_campground_playground-03.JPG

One of the things that the 2013 flood of the Bow River teaches us is that we cannot predict when a flood will occur nor how big it will be, so in order to minimize damage and casualties we need to be prepared. Some ways of preparing include:

- Mapping flood plains and not building within them
- Building dykes or dams where necessary
- Monitoring the winter snowpack, the weather, and stream discharges
- Creating emergency plans
- Educating the public on how to prepare for and respond to the threat of flooding

Exercise 14.5 Flood Probability of the Bow River

The graph below shows the highest discharge per year between 1915 and 2014 of the Bow River in Calgary. Using this data set, we can calculate the recurrence interval (Ri) for any particular flood magnitude using the equation: Ri = (n+1)/r (where n is the number of floods in the record being considered, and r is the rank of the particular flood). There are a few years missing in this record, so the actual number of floods is 95.

Source: Steven Earle (2015) CC BY 4.0 view source https://opentextbc.ca/geology/, from data at Water Surveys of Canada, Environment Canada, http://wateroffice.ec.gc.ca/search/searchDownload_e.html

The largest flood recorded along the Bow River over that period of time was the one in 2013, reaching a peak flow rate of 1,840 m^3/s on June 21. Ri for this flood is (95+1)/1 = 96 years. The probability of such a flood in any future year is 1/Ri x 100%, which is 1%. The fifth largest flood was just a few years earlier in 2005, at 791 m^3/s. Ri for this flood is (95+1)/5 = 19.2 years. The recurrence probability of a flood of this magnitude is thus 5%.

1. Calculate the recurrence interval for the second largest flood (1932, 1,520 m^3/s).

2. What is the probability that a flood of 1,520 m^3/s will happen next year?

3. Examine the 100-year trend for floods along the Bow River. If you ignore the major floods (the labelled ones), what is the general trend of peak discharges over this time?

Chapter Summary

The topics covered in this chapter can be summarized as follows:

14.1 The Hydrological Cycle

Water is stored in the oceans, glacial ice, the ground, lakes, rivers, and the atmosphere. Its movement is powered by solar energy and gravity.

14.2 Drainage Basins

All of the precipitation that falls within a drainage basin flows into the stream that drains that area. Stream drainage patterns are determined by the type of rock within the basin. Over geological time, streams change the landscape that they flow within, and eventually they become graded, meaning their profile becomes a smooth curve. A stream can lose that gradation if there is renewed uplift or if their base level changes for some other reason such as construction of a dam downstream.

14.3 Stream Erosion and Deposition

The processes of erosion and deposition of particles within streams are primarily driven by the velocity of the stream water. Erosion and deposition of different-sized particles can happen simultaneously in a stream. Some particles are moved along the bottom of a river while others are carried in suspension. It takes a greater velocity of water to erode a particle from a stream bed than it does to keep it in suspension. Ions are also transported in solution. When a stream rises and then occupies its flood plain, the velocity of water over the flood plains slows and natural levees form along the edges of the stream channel.

14.4 Stream Types

Youthful streams in steep areas erode most rapidly downwards, and they tend to have steep, rocky, and relatively straight channels. Where sediment-rich streams empty into areas with lower gradients, braided streams can form. Meandering streams are common in areas with even lower gradients where silt and sand are the dominant sediments. Meandering streams erode the walls of their channels more rapidly than the channel base. Deltas form where streams flow into standing water.

14.5 Flooding

Most streams in Canada have their highest discharge rates in spring and early summer, although the highest discharge in many of BC's coastal streams is in the winter. Floods happen when a stream rises high enough to spill over its banks and spread across its flood plain. Some of the more significant floods in Canada include the Fraser River flood of 1948, the Saguenay River flood of 1996, the Red River flood of 1997, and the Alberta floods of 2013. We can estimate the probability of a specific flood level based on the record of past floods, and we can take steps to minimize the impacts of flooding such as building floodways to divert excess water and not building in flood-prone areas.

Questions for Review

1. What is the proportion of liquid *fresh* water on Earth expressed as a percentage of all water on Earth?
2. What percentage of this fresh water is groundwater?
3. What type of rock, and what processes, can lead to the formation of a trellis drainage pattern?
4. Why do many of the streams in the southwestern part of Vancouver Island empty into the ocean over waterfalls?
5. Where would you expect to find the fastest water flow along a straight stretch of a stream?
6. Sand grains can be moved by traction and saltation. What minimum stream velocity is required to move 1 mm sand grains?
7. If the flow velocity of a stream is 1 cm/s, what sizes of particles can be eroded, what sizes can be transported if they are already in suspension, and what sizes of particles cannot be moved at all?
8. Under what circumstances might a braided stream develop?
9. How would the gradient of a stream be affected if a meander is bypassed?
10. The elevation of the Fraser River at Hope is 41 m. From there it flows approximately 147 km to the sea. What is the average gradient of the river (m/km) over this distance?
11. How do BC's coastal streams differ from most of the rest of the streams in Canada in terms of their annual flow patterns? Why?
12. Why do most serious floods in Canada happen in late May, June, or early July?
13. There is a 65-year record of peak annual discharges along the Ashnola River near Princeton, BC. During this time, the second highest discharge recorded was 175 m^3/s. Based on this information, what is the recurrence interval (Ri) for this discharge level, and what is the probability that there will be a similar peak discharge next year?

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-SA-NC 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 15. Mass Wasting

Adapted by Joyce M. McBeth, University of Saskatchewan
from Physical Geology by Steven Earle

Learning Objectives

After carefully reading this chapter, completing the exercises within it, and answering the questions at the end, you should be able to:

- Explain how slope stability is related to slope angle
- Summarize some of the factors that influence the strength of materials on slopes, including the type of rock, the presence and orientation of planes of weakness such as bedding planes or fractures, the type of unconsolidated material upon the slope, and the effects of water flowing upon or within the slope material
- Explain what types of events can trigger mass wasting
- Summarize the types of motion that can happen during mass wasting
- Describe the difference between a translational and a rotational slope failure
- Describe the main types of mass wasting — creep, slump, slide, fall, and debris or mud flow — in terms of the types of materials involved, the type of motion, and the likely rates of motion
- Explain what steps we can take to delay mass wasting, and why we cannot prevent it permanently
- Describe some of the measures that can be taken to mitigate the risks associated with mass wasting

Mass wasting is the failure and downslope movement of rock or unconsolidated materials in response to gravity. Mass wasting is an agent of erosion. The term “landslide” is used synonymously with the term mass wasting, but they are not the same thing. Some people reserve the term “landslide” for relatively rapid slope failures, while others do not. Due to this ambiguity, we avoid the use of the term “landslide” in this textbook.

The Hope Slide: a Historic Canadian Example of a Mass Wasting Event

Figure 15.1 The site of the 1965 Hope Slide, photographed in 2014. The initial failure is thought to have taken place along foliation planes in the rock and a sill. *Source: Steven Earle (2015) CC BY 4.0. view source https://opentextbc.ca/geology/*

Early in the morning of January 9, 1965, 47 million cubic metres of rock broke away from the steep upper slopes of Johnson Peak (16 km southeast of Hope, B.C.) and tumbled 2,000 m down the mountain, gouging out the contents of a small lake at the bottom, and continuing a few hundred metres up the other side of the valley (Figure 15.1). Four people were killed who had been stopped on the highway by a snow avalanche. Many more people might have become victims, except that a Greyhound bus driver, en route to Vancouver, turned his bus around upon seeing the avalanche. The rock failed along foliation planes of the metamorphic rock on Johnson Peak, in an area that had been eroded into a steep slope by glacial ice. There is no evidence that it was triggered by any specific event, and there was no warning that it was about to happen. Even if there had been warning, nothing could have been done to prevent it. There are hundreds of similar sites throughout mountainous regions of British Columbia and elsewhere in Canada where large mass wasting events could occur.

What can we learn from the Hope Slide? In general, we cannot prevent most mass wasting events, and significant effort is required if an event is to be predicted with any level of certainty. Understanding the geology is critical to understanding mass wasting. Although slope failures are inevitable in a region with steep slopes, larger slope failures happen less frequently than smaller ones. The consequences of a large mass wasting event also vary depending on the downslope conditions, such as the presence of people, buildings, roads, or fish-bearing streams.

An important reason for learning about mass wasting is to understand the nature of how and why materials fail in mass wasting events. If we understand this better, we can use this knowledge to help minimize the risks from similar events in the future.

15.1 Factors That Control Slope Stability

Slope Angle and the Forces Acting Upon A Slope

A block of rock situated on a rock slope is pulled by gravity toward Earth's centre (vertically down, Figure 15.2). We can split the vertical gravitational force into two components (vectors) relative to the slope: one pulling the block down parallel to the slope (the **shear force**), and the other pulling the block directly into (i.e., perpendicular) to the slope (the **normal force**).

The shear force pulls the block down the slope, but the block does not move unless the shear force overcomes (is greater than) the friction between the block and the slope. This friction holding the block on the slope may be quite weak if the block has split away from the main body of rock, or may be very strong if the block is still connected to the other rock on the slope. The strength of the relationship between the block and the slope is called the **shear strength**. For example, in Figure 15.2a, the shear strength is greater than the shear force, so the block will not move. In Figure 15.2b the slope is steeper, and the shear force is approximately equal to the shear strength. The block may or may not move under these circumstances. In Figure 15.2c, the slope is steeper still, so the shear force is considerably greater than the shear strength, and the block will move.

Figure 15.2 Differences in the shear and normal components of the gravitational force on slopes with differing steepness. The total gravitational force is the same in all three cases. In (a) the shear force (red line aligned with slope) is substantially less than the shear strength (green arrow), so the block is stable. In (b) the shear force and shear strength are nearly equal, so the block may or may not move. In (c) the shear force is greater than the shear strength, so the block will move. *Source: Steven Earle (2015) CC BY 4.0. view source https://opentextbc.ca/geology/*

Slopes are created by uplift in the Earth's crust and modified by erosion. In areas with relatively recent uplift (such as most of British Columbia and the western part of Alberta in Canada), slopes tend to be quite steep. This is especially true where glaciation has taken place because glaciers in mountainous terrain create steep-sided U-shaped valleys. In areas without recent uplift (such as central Canada), slopes are less steep because they have been subjected to erosion, including mass

wasting, for long periods of time. Note that mass wasting can happen even on relatively gentle slopes if the shear stress acting on the materials is greater than the materials' shear strength.

Slope Strength

The strength of the materials on slopes can vary widely. Solid rocks tend to be strong, but rock strength varies widely, so this is not always the case. If we consider just the strength of the rocks and ignore issues such as fracturing and layering, then most crystalline rocks (e.g., granite, basalt, or gneiss) are very strong, while some metamorphic rocks (e.g., schist) are only moderately strong. Sedimentary rocks have variable strength. Dolostone and some limestone are strong, most sandstone and conglomerate are moderately strong, and some sandstone and all mudstones are quite weak.

Fractures, metamorphic foliation (excluding gneissosity and banding), or bedding can significantly reduce the strength of rock. In the context of mass wasting, this is most critical if the planes of weakness are parallel to the slope and least critical if they are perpendicular to the slope. This is illustrated in Figure 15.3. At locations A and B the bedding is nearly perpendicular to the slope and the layers of rock are relatively stable. At location D the bedding is nearly parallel to the slope and the layers of rock are relatively unstable. At location C the bedding is nearly horizontal, and the stability is intermediate between the two extremes.

Figure 15.3 Relative stability of slopes. The stability is as a function of the orientation of planes of weakness (in this case bedding planes) relative to the slope orientations. *Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Internal variations in the composition and structure of rocks can significantly affect their strength. Schist, for example, may have layers that are rich in sheet silicates (micas) and these will tend to form weak layers. Some minerals tend to be more susceptible to weathering than others, and the weathered products are commonly quite weak (e.g., clay formed from feldspar). The side of Johnson Peak that failed in 1965 (Hope Slide) is made up of chlorite schist (metamorphosed sea-floor basalt) that has feldspar-bearing sills within it. The foliation and the sills are parallel to the steep slope. The schist is relatively weak to begin with, and the feldspar in the sills, which has been altered to clay, makes it even weaker.

Unconsolidated sediments are generally weaker than sedimentary rocks because they are not cemented and, in most cases, have not been significantly compressed by overlying materials. Unconsolidated sediments can still bind together, and the strength of that binding is called cohesion. A cohesive sediment binds together strongly and if you picked it up with a shovel it would stick together in a lump (e.g., sand mixed with clay, clay). A sediment that is not very cohesive is weakly bound and would probably fall apart if you picked it up with a shovel (e.g., sand, silt). The deposits that make up the cliffs at Point Grey, Vancouver, B.C. include sand, silt, and clay, overlain by sand. The finer deposits at Point Grey are relatively cohesive (they maintain a steep slope, Figure 15.4 left). The overlying sand is not very cohesive (relatively weak) and has a shallower slope because there are many slope failures in the sand deposit.

In contrast to poorly cohesive sediment deposits, glacial till can be as strong as some sedimentary rock. Glacial till is typically a mixture of clay, silt, sand, gravel, and larger clasts and forms and is compressed beneath tens to thousands of metres of glacial ice (Figure 15.4, right).

Figure 15.4 Left: Glacial outwash deposits at Point Grey, Vancouver, B.C. The dark lower layer is made up of sand, silt, and clay. The light upper layer is well-sorted sand, which has experienced slope failure and formed a cone of talus. Right: Glacial till on Quadra Island, B.C. The till is strong enough to have formed a near-vertical slope. *Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Apart from the type of material on a slope, the amount of water that the material contains is the most important factor controlling its strength. This is especially true for unconsolidated materials (e.g., Figure 15.4), but it also applies to bodies of rock. Granular sediments, like the sand at Point Grey, have lots of pore spaces between the grains. These spaces may be completely dry (filled only with air), moist (some spaces are water filled), or completely saturated (Figure 15.5).

Figure 15.5 Depiction of dry, moist, and saturated sand. *Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Unconsolidated sediments tend to be strongest when they are moist because the small amounts of water at grain boundaries holds the grains together due to surface tension. Surface tension is the tension at the surface of a fluid that allows the liquid to resist an external force. Liquids always tend to acquire the lowest surface area possible; this happens because molecules at the surface of the fluid are attracted to the molecules below the surface). This is the property of liquid water that allows insects to walk over it. Dry sediments are held together only by the friction between grains, and if they are well sorted or well rounded, or both, this cohesion is weak, due to minimal grain contact. Saturated sediments tend to be the weakest of all because the water pushes the grains apart, decreasing friction between grains. Water will also reduce the strength of solid rock, if the rock has porosity, fractures, bedding planes, and/or clay-bearing zones, especially when the rock is saturated with water (saturated conditions).

Water pressure is an important factor in slope failure. As you move deeper in saturated sediment, the pressure of the water goes up due to gravity acting on the column of water above it; this pressure is called hydrostatic pressure. The greater the depth below the surface of the water table (the point where the rock or sediments are saturated), the greater the water pressure acting on the materials. Holes are often drilled into rocks in road cuts to allow water to drain and relieve this water pressure. One of the hypotheses advanced to explain the 1965 Hope Slide is that cold conditions that winter caused small springs in the lower part of the slope to freeze, preventing water from flowing out. It is possible that water pressure gradually built up within the slope, weakening the rock mass to the extent that the shear strength was no longer greater than the shear force.

Water also has an interesting effect on clay-bearing materials. All clay minerals will absorb a small quantity of water, which reduces the strength of the clay. The smectite clays (such as the bentonite used in cat litter) can absorb a lot of water, and this water pushes the clay sheets apart at a molecular level, which makes the clay swell. Smectite that has expanded in this way has almost no strength; it is extremely slippery. Thus, slopes containing smectite clay are more likely to experience slope failure when they are saturated.

Water can increase the mass of the material on a slope, because the mass of the water is a component of the overall mass of the slope material. This increases the gravitational force pulling the slope materials down. A water saturated body of sediment with 25% porosity weighs approximately 13% more than it does when it is completely dry, so the gravitational shear force is also 13% higher. In the

situation shown in Figure 15.2b, a 13% increase in the shear force is enough to overcome the shear strength, and the block would move down the slope.

Exercise 15.1 Sand and Water

Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/

If you have ever been to the beach, you already know that sand behaves differently when it is dry than it does when it is wet. The following experiment will demonstrate the strength of sand when it is dry, moist and saturated.

Find approximately half a cup of clean, dry sand (or get some wet sand and dry it out), and then pour it from your hand onto a piece of paper. You should be able to make a cone-shaped pile that has a slope of ~30°. If you pour more sand onto the pile, it will get bigger, but the slope should remain the same.

Now add some water to the sand so that it is moist. One way to do this is to add enough water to saturate the sand, then let the water drain away for a minute. You should be able to form this moist sand into a steep pile (with slopes of ~80°).

Finally, put some sand into a cup and fill the cup with water so the sand is just covered. Swirl it around so that the sand remains in suspension, and then quickly tip it out onto a flat surface. It should spread out over a wide area, forming a pile with a slope of only a few degrees.

Triggers of Mass Wasting

In the previous section, we discussed the shear force and the shear strength of materials on slopes, and factors that can decrease the shear strength. Shear force is primarily related to slope angle, and once a slope angle is set, the shear force is constant. But shear strength can change quickly for a variety of reasons. Events that lead to a rapid decrease in shear strength are triggers for mass wasting.

An increase in water content is the most common trigger of mass wasting. This can result from rapid melting of snow or ice, heavy rain, or other events that change the pattern of water flow on and through the material. Rapid melting can be caused by a dramatic increase in temperature (e.g., in spring or early summer), or by a volcanic eruption. Heavy rains are typically related to storms. An example of a major slope failure caused by an increase in water content was the Oso landslide that occurred in Washington State, USA in 2014 (Figure 15.6). The flow buried the community of Steelhead Haven and killed 43 people.

Figure 15.6 | The Oso landslide, a flow that occurred in Washington State, USA 22 March 2014. Source: Matthew Sissel (2014) Public Domain. View source. https://www.dvidshub.net/image/1209679/oso-mudslide

Changes in water flow patterns can be caused by earthquakes, dammed streams from previous slope, or human structures that interfere with runoff (e.g., buildings, roads, or parking lots). An example of this is a deadly 2005 debris flow that occurred in North Vancouver, B.C. This slope failure took place in an area where there had been previous slope failures. A report written in 1980 recommended that the municipal authorities and residents take steps to address surface and slope drainage issues. Unfortunately, little was done to improve the situation or to take steps that could have prevented the

2005 slope failure. The failure happened during a rainy period but was likely triggered by excess runoff related to the roads at the top of this slope and by landscape features including addition of fill to backyards in the area above the failure.

In some cases, a decrease in water content can lead to failure. This is most common with clean sand deposits (e.g., the upper layer in Figure 15.4 (left)), which loses some of its strength when most of the water around the grains is removed (i.e., as the sand water content drops the surface tension decreases).

Freezing and thawing can also trigger some forms of mass wasting e.g., thawing can release a block of rock that was frozen to a slope by a film of ice.

One other process that can weaken a body of rock or sediment is shaking. The most obvious source of shaking is an earthquake. Shaking from highway traffic, construction, or mining can also weaken rock. Several deadly mass wasting events (including snow avalanches) were triggered by the M7.8 earthquake in Nepal in April 2015.

15.2 Classification of Mass Wasting

There are three criteria used to classify slope failures:

- The type of material that failed (e.g., bedrock or unconsolidated sediment),
- The mechanism of the failure (how the material moved as it failed),
- The rate of movement (how quickly the material moved).

The mechanism of the failure is the most important characteristic of a slope failure. There are three main mechanisms of slope failure:

- **Fall**: in falls, the material falls through the air, either vertically or nearly vertically,
- **Slide**: in slides, the material moves as a coherent mass along a sloping surface (with little to no internal motion within the mass), and
- **Flow**: in flows the failing material has internal motion (like a fluid) as it moves.

While we do not classify slope failures by the shape of the rupture surface, it is nevertheless an important feature used to describe mass wasting processes. The **rupture surface** is the boundary between the slope and the sliding material. Slope failures with curved rupture surfaces are called **rotational slope failures**, and slope failures with planar rupture surfaces are called **translational slope failures.**

Unfortunately, classifying slope failure is not as simple as the classification scheme above suggests**.** Many slope failures involve more than one type of motion, and often it is not easy to tell how the material moved. The types of slope failure that are covered in this chapter are summarized in Table 15.1.

Failure Type	Type of Material	Type of Motion	Rate of Motion
Rock fall	Rock fragments	Vertical or near-vertical fall (plus bouncing in many cases)	Very fast (>10s m/s)
Rock slide	A large rock body	Motion as a unit along a planar surface (translational sliding)	Typically very slow (mm/y to cm/y), but some can be faster
Rock avalanche	A large rock body that slides and then breaks into small fragments	Flow (at high speeds, the mass of rock fragments is suspended on a cushion of air)	Very fast (>10s m/s)
Creep or solifluction	Soil or other overburden; in some cases, mixed with ice	Flow (although sliding motion may also occur)	Very slow (mm/y to cm/y)
Slump	Thick deposits (m to 10s of m) of unconsolidated sediment	Motion as a unit along a curved surface (rotational sliding)	Slow (cm/y to m/y)
Mudflow	Loose sediment with a significant component of silt and clay	Flow (a mixture of sediment and water moves down a channel)	Moderate to fast (cm/s to m/s)
Debris flow	Sand, gravel, and larger fragments	Flow (similar to a mudflow, but typically faster)	Fast (m/s)

Table 15.1 Classification of slope failures based on type of material and motion. *Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Rock Fall

Rock fragments can break off relatively easily from steep bedrock slopes, most commonly due to frost-wedging. If you ever hike along a steep mountain trail on a cool morning, you will probably hear the occasional fall of rock fragments onto a **talus slope.** Water in the cracks in the rock freezes and expands overnight, and this can break the rock apart. When the ice thaws in the morning sun, some of these broken fragments will fall to the slope below (Figure 15.7). Talus slopes form from this fallen rock, forming slopes at the angle of repose below high rock walls (Figure 15. 8, left)**.** They are also known as scree slopes.

A typical talus slope, near Keremeos in southern BC, is shown in Figure 15.8. In December 2014, a large block of rock split away from a cliff in this same area. It broke into smaller pieces that tumbled down the slope and crashed into the road, smashing the concrete barriers and gouging out large parts of the pavement. Luckily no one was hurt.

Figure 15.7 The contribution of freeze-thaw to a rock fall. *Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Figure 15.8 Left: A talus slope near Keremeos, B.C., formed by rock falls from the cliffs above. Right: The results of a rock fall onto a highway west of Keremeos in December 2014. *Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Rock Slide

A rock slide is a large body of rock that is sliding along a sloping surface. In most cases, the movement is parallel to a fracture, bedding plane, or metamorphic foliation plane; thus, most rock slides are translational slope failures.

The speed of the movement can range from very slow to moderately fast. The word **sackung** describes the very slow motion of a block of rock (mm/y to cm/y) on a slope. A good example is the Downie Slide, a translational slide north of Revelstoke, BC, which is shown in Figure 15.9. In this case, a massive body of rock is very slowly sliding down a steep slope along a plane of weakness that is caused by the foliation in the rock. The foliation is approximately parallel to the slope.

The Downie Slide was recognized prior to the construction of the Revelstoke Dam, and was moving very slowly at the time of dam construction (a few cm/year). Geological engineers were concerned that the presence of water in the reservoir (visible in Figure 15.9) could further weaken the plane of failure, leading to an acceleration of the motion. The result would have been a catastrophic failure into the reservoir that would have sent a wall of water over the dam and into the community of Revelstoke.

Figure 15.9 The Downie Slide, a sackung, on the shore of the Revelstoke Reservoir (above the Revelstoke Dam). The head scarp is visible at the top and a side-scarp along the left side. Source: Joyce McBeth (2018) CC BY 4.0, image © 2018 Google Earth, Data Google CNES / Airbus Data LDEO-Columbia, NSF, NOAA Data SIO, NOAA, U.S. Navy, NGA, GEBCO DigitalGlobe Landsat / Copernicus Province of BC.

During the construction of the dam, the engineers tunnelled into the rock at the base of the slide and drilled hundreds of drainage holes upward into the plane of failure. This allowed water to drain out more efficiently so that the hydrostatic pressure was decreased, which decreased the rate of movement of the sliding block. BC Hydro monitors this site continuously. The slide block is currently moving more slowly than it was prior to the construction of the dam.

In the summer of 2008, a large block of rock slid rapidly from a steep slope above Highway 99 near Porteau Cove, BC (between Horseshoe Bay and Squamish). The block crashed into the highway and adjacent railway and broke into many pieces, and the highway was closed for several days. Engineers and geoscientists took steps to stabilize the slope to decrease the risk of future rock falls. Rock bolts (long metal rods) were installed to anchor the blocks of rocks and prevent them falling Drainage holes were installed to drain water from the slope and decrease the water pressure. As shown in Figure 15.10, the rock along the slope is fractured parallel to the slope, and this almost certainly contributed

to the failure. However, it is not actually known what triggered this event as the weather was dry and warm during the preceding weeks, and there was no significant earthquake in the region.

Figure 15.10 Site of the 2008 rock slide at Porteau Cove. Notice the prominent fracture set parallel to the surface of the slope. The slope has been stabilized with rock bolts (top arrow) and holes have been drilled into the rock to improve drainage (tube from drainage hole indicated with bottom arrow). Risk to passing vehicles from rock fall has been reduced by hanging mesh curtains (background), which secures loose material to the slope. *Source: Joyce McBeth (2018) CC BY 4.0 after Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Rock Avalanche

If a rock slides and then starts moving quickly (m/s), the rock is likely to break into many small pieces. At this point it is considered to be a **rock avalanche**, in which the large and small fragments of rock move in a fluid manner supported by a cushion of air within and beneath the moving mass. The 1965 Hope Slide (Figure 15.1) was a rock avalanche, as was the famous 1903 Frank Slide in southwestern Alberta. The 2010 slide at Mt. Meager (west of Lillooet) was also a rock avalanche and rivals the Hope Slide as the largest slope failure in Canada during historical times (Figure 15.11).

Figure 15.11 The 2010 Mt. Meager landslide, showing where the slide originated (arrow, 4 km upstream). It then raced down a steep narrow valley and out into the wider valley in the foreground. Source: Mika McKinnon (2011) CC BY-SA-NC, View source http://www.geomika.com/blog/2011/01/05/the-trouble-with-landslides/

Creep or Solifluction

The very slow — mm/y to cm/y — movement of soil or other unconsolidated material down slope is known as **creep**. Creep, which normally only affects the upper several centimetres of loose material, is typically a type of very slow flow. In some cases, sliding may take place too.

Creep can be facilitated by freezing and thawing because, as shown in Figure 15.12, particles are lifted perpendicular to the surface by the growth of ice crystals within the soil, and then move downwards vertically due to gravity when the ice melts. The same effect can be produced by frequent wetting and drying of the soil. In cold environments, **solifluction** is a more intense form of freeze-thaw-triggered creep.

Creep is most noticeable on moderate-to-steep slopes where trees or fence posts are consistently leaning in a downhill direction. In the case of trees, they try to correct their lean by growing upright, and this leads to a curved lower trunk known as a "pistol butt." Creep can also generate terracettes, horizontal and repeating ridges on slopes (Figure 15.13). Historically, people thought terracettes formed where livestock or wild animals regularly travelled along slopes. While animals can accentuate terracettes, the primary reason terracettes form is creep.

Figure 15.12 A depiction of the contribution of freeze-thaw to creep. The blue arrows represent uplift caused by freezing in the wet soil underneath, while the red arrows represent depression by gravity during thawing. The uplift is perpendicular to the slope, while the drop is vertical. *Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Slump

A slide is a mass movement where the material moves as a coherent mass. A **slump** is a type of slide that takes place within thick unconsolidated deposits (typically thicker than 10 m). Slumps involve movement along one or more curved failure surfaces and are thus rotational slope failures. Slumps have downward motion near the top and outward motion toward the bottom (Figure 15.14). They are typically caused by high water pressure within these materials on a steep slope.

Figure 15.13 Evidence of creep-generated terracettes on the Peace River hills in northeastern B.C. *Source: Joyce McBeth (2018) CC BY 4.0.*

An example of a slump in the Lethbridge area of Alberta is shown in Figure 15.15. This feature has likely been active for many decades and moves a little more whenever there are heavy spring rains and snowmelt runoff. The toe of the slump is being eroded by the small stream at the bottom. The erosion contributes to continued slumping. The basal material (material at the toe of the slope) forms the support for the overlying mass of material in the slope and if this support is eroded away slumping will likely continue.

Figure 15.14 The motion of unconsolidated sediments in an area of slumping. *Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Figure 15.15 A slump along the banks of a small coulee near Lethbridge, Alberta. The main head-scarp is clearly visible at the top, and a second smaller one is visible about a quarter of the way down the slope. The toe of the slump is being eroded by the seasonal stream that created the coulee. *Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Mudflows and Debris Flows

As you saw in Exercise 15.1, when a mass of sediment becomes completely saturated with water, the mass loses strength, to the extent that the grains may be pushed apart and may flow, even on a gentle slope. This can happen during rapid spring snowmelt or heavy rains and is also relatively common during volcanic eruptions because of rapid melting of snow and ice. If the material involved is primarily sand-sized or smaller, it is known as a mudflow, such as the one shown in Figure 15.16. A mudflow or debris flow on a volcano or during a volcanic eruption is called a *lahar*.

If the material involved is gravel sized or larger, it is known as a debris flow. Since it takes more gravitational force to overcome friction and move larger particles, debris flows typically form in areas with steeper slopes and higher water pressure. In many cases, a debris flow takes place within a steep stream channel and is triggered by the collapse of bank material into the stream. This may create a temporary dam followed by a major flow of water and debris when the dam finally bursts. This is the situation that led to the fatal debris flow at Johnsons Landing, BC, in 2012.

A typical west-coast debris flow is shown in Figure 15.17. This event took place in November 2006 in response to very heavy rainfall. There was enough energy in the flow to move large boulders and to knock over large trees.

Figure 15.16 A slump (left) and an associated mudflow (centre) at the same location as Figure 15.15, near Lethbridge, Alberta. *Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Figure 15.17 The lower part of debris flow within a steep stream channel near Buttle Lake, B.C., in November 2006. Note the trees along the edges of the stream that have been damaged by the rocks in the debris flow. *Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Exercise 15.2 Classifying Slope Failures

These four photos show some of the different types of slope failures described above. Try to identify each and provide some criteria to support your choice.

Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/

15.3 Preventing, Delaying, Monitoring, and Mitigating Mass Wasting

We cannot prevent mass wasting, however, in many situations there are actions we can take to reduce or mitigate the damaging effects of mass wasting on people and infrastructure. Where we can neither delay nor mitigate mass wasting, we may consider trying to initiate the slope failure in a controlled manner. In areas prone to mass wasting that cannot be controlled or mitigated, we can minimize risk by not building in these areas at all.

Preventing and Delaying Mass Wasting

It is comforting to think that we can prevent some effects of mass wasting by mechanical means. For example, the rock bolts in the road cut at Porteau Cove on the Sea-toSky highway in BC (Figure 15.10) or the drill holes used to drain water out of the slope at the Downie Slide (Figure 15.9), or the building of physical barriers, such as retaining walls along highway roadcuts. These preventative measures are not permanent though, they are subject to degradation over time. The rock bolts in the road cut at Porteau Cove will slowly start to corrode after a few years, and within a few decades many of them will begin to lose their strength. Unless they are replaced, they will no longer support the slope. Likewise, drainage holes at the Downie Slide will eventually become plugged with sediment and chemical precipitates, and unless they are periodically unplugged, their effectiveness will decrease. Eventually, unless new holes are drilled, the drainage will be compromised, and the slide will start to move again. This is why careful slope monitoring by geological and geotechnical engineers is important at major mass wasting sites such as the Downie Slide and along the Sea-to-Sky highway. Our efforts to control mass wasting are only as good as our efforts to maintain the preventive measures.

Delaying mass wasting is a worthy endeavour because during the time that the measures are still effective, they can save lives and reduce damage to property and infrastructure such as homes and roads. But we must be careful to avoid activities that could make mass wasting more likely. One of the most common anthropogenic causes of mass wasting is road construction, and this applies both to remote gravel roads built for forestry and mining, and large urban and regional highways.

Figure 15.18 An example of a road constructed by cutting into a steep slope and the use of the cut material as fill. *Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Road construction is a potential problem for two reasons. First, creating a flat road surface on a slope inevitably involves creating a cut bank that is steeper than the original slope. This might also involve creating a filled bank that is both steeper and weaker than the original slope (Figure 15.18). Second, roadways typically cut across natural drainage features, and unless great care is taken to reroute the runoff water, oversaturation of slope material can occur, contributing to mass wasting.

Apart from saturation and water pressure considerations, engineers building roads and other infrastructure on bedrock slopes have to carefully consider the geology, and especially any weaknesses or discontinuities in the rock related to bedding, fracturing, or foliation. If possible,

situations like that at Porteau Cove (Figure 15.10) should be avoided — by building somewhere else — rather than trying to stitch the slope back together with rock bolts.

It is widely believed that construction of buildings above steep slopes can contribute to the instability of the slope. This is likely true, but probably not because of the weight of the building. As you will determine by completing Exercise 15.3 below, a typical house is not heavier than the excavated ground that was removed to build the house. A more likely contributor to instability of the slopes below buildings is changes to the water drainage and to the saturation of the slope (by watering gardens, for example).

Exercise 15.3 How Much Does A House Weigh and Can It Contribute to Slope Failure?

It is commonly believed that building a house (or some other building) at the top of a slope will add a lot of extra weight to the slope, which could contribute to slope failure. But what does a house weigh compared to the material removed when you build it? A typical 150 m^2 (approximately 1,600 ft^2) wood-frame house with a basement and a concrete foundation weighs about 145 t (metric tonnes). But most houses are built on foundations that are excavated into the ground. This involves digging a hole and taking that material away, so we need to subtract what that excavated material weighs. Assuming our 150 m^2 house required an excavation that was 15 m by 11 m by 1 m deep, which is 165 m^3 of material. Unconsolidated sediments have densities ranging from about 0.8 to 1.7 t per m^3.

For this exercise, consider a sand with a dry density of 1.2 t per m^3 for this calculation. Calculate the weight of the materials that were removed and compare that with the weight of the house and its foundation.

If you are thinking that building a bigger building is going to add more weight, consider that bigger buildings need bigger and deeper excavations, and in many cases the excavations may be into solid rock, which is denser than surficial materials.

Consider how a building might change the drainage on a slope. Water can be collected by roofs, go into downspouts, and form concentrated flows that are directed onto or into the slope. Likewise, drainage from nearby access roads, lawn irrigation, leaking pools, and septic systems can all alter the surface and groundwater flow in a slope.

Soil excavated from a basement. Source: Steven Earle (2015) CC BY 4.0. View source *https://opentextbc.ca/geology/*

Monitoring Mass Wasting

Warning systems are helpful in some areas where there is a risk of mass wasting. They let us know if conditions have changed at a known slide area, or if a rapid failure, such as a debris flow, is on its way downslope. The Downie Slide above the Revelstoke Reservoir is continuously-monitored with a range of devices, such as inclinometers (slope-change detectors), bore-hole motion sensors, and GPS survey instruments. A simple mechanical device for monitoring the nearby Checkerboard Slide (which is also above the Revelstoke Reservoir) is shown in Figure 15.19. Both of these slides are very slow-moving, but it is important to be able to detect changes in their rates of motion. A rapid failure would result in large bodies of rock plunging into the reservoir and sending a wall of water over the Revelstoke Dam, potentially destroying the nearby town of Revelstoke.

Mt. Rainier, a glacier-covered volcano in Washington State (15.20), could produce massive mudflows or debris flows (lahars) with or without a volcanic eruption. Over 100,000 people in the Tacoma, Puyallup, and Sumner areas are at risk because they currently reside on deposits from past lahars and future lahars would likely also follow these paths (Figure 15.21). In 1998, a network of acoustic monitors was established around Mt. Rainier. The monitors are embedded in the ground adjacent to expected lahar paths. These monitors will provide warnings to emergency officials in the event of a lahar. When a lahar is detected, the residents of the area will have between 40 minutes and three hours to get to safe ground.

Figure 15.19 Part of a motion-monitoring device at the Checkerboard Slide near Revelstoke, BC. The lower end of the cable (extending out from the top of the device to the right) is attached to a block of rock that is unstable. Any incremental motion of this block will move the cable, which will be detectable by this device. *Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Figure 15.20 Mt. Rainier from Seattle, WA, USA. Source: CC-BY-NC-ND 2.0. View source https://flic.kr/p/8yeiGu

Figure 15.21 Major pathways of Mt Rainier lahars over the past 10,000 years, Washington State, USA. Source: USGS (2005) Public Domain view source https://volcanoes.usgs.gov/volcanoes/mount_rainier/hazard_lahars.html, modified from Driedger et al (2005) Public Domain. View source https://pubs.er.usgs.gov/publication/gip19

Mitigating the Impacts of Mass Wasting

In situations where we cannot predict, prevent, or delay mass-wasting hazards, some effective measures can be taken to minimize the associated risk. For example, many highways in BC and western Alberta have avalanche shelters like the one shown in Figure 15.22. In some parts of the world, similar structures have been built to protect infrastructure from other types of mass wasting.

Figure 15.22 | A snow avalanche shelter on the Coquihalla Highway (bottom centre of the image). The expected path of the avalanche is the steep and treeless slope above. *Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Debris flows are inevitable, unpreventable, and unpredictable in many parts of BC, but nowhere more so than along the Sea-to-Sky Highway between Horseshoe Bay and Squamish. The results have been deadly and expensive many times in the past. It would be very expensive to develop a new route in this region, so provincial authorities have taken steps to protect residents, and traffic on the highway and railway. Debris flow defensive structures have been constructed in several drainage basins, as shown in Figure 15.23. One strategy is to allow the debris flow to flow quickly through to the ocean along a smooth channel. Another is to capture the debris within a constructed basin that allows the excess water to continue through.

Figure 15.23 | Two strategies for mitigating debris flows on the Sea-to-Sky Highway. Left: A concrete –lined channel on Alberta Creek allows debris to flow quickly through to the ocean. Right: A debris flow catchment basin on Charles Creek. In 2010, a debris flow filled the basin to the level of the dotted white line. *Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Finally, in situations where we cannot do anything to delay, predict, contain, or mitigate slope failures, the responsible and ethical thing to do is to avoid building in or using the risky area.

Sometimes this may require relocating a community after discovering a previously-unidentified risk. There is a famous example of this in BC at a site known as Garibaldi, 25 km south of Whistler. In the early 1980s the village of Garibaldi had a population of about 100, with construction underway on some new homes, and plans for many more. In the months that followed the deadly 1980 eruption of Mt. St. Helens in Washington State, the BC Ministry of Transportation commissioned a geological study to assess risks along their highways. The study revealed that a steep cliff known as The Barrier (Figure 15.24) had collapsed in 1855, leading to a large rock avalanche, and that it was likely to collapse again unpredictably, putting the village of Garibaldi at extreme risk. In an ensuing court case, it was ruled that the Garibaldi site was not a safe place for people to live. Those who already had homes there were compensated, and everyone was ordered to leave.

Figure 15.24 | The Barrier, south of Whistler, BC, was the site of a huge rock avalanche in 1855, which extended from the cliff visible here 4 km down the valley and across the current location of the Sea-to-Sky Highway and the Cheakamus River. *Source: Steven Earle (2015) CC BY 4.0. View source https://opentextbc.ca/geology/*

Chapter Summary

The topics covered in this chapter can be summarized as follows:

15.1 Factors That Control Stability on Slopes

Slope stability is controlled by the slope angle and the strength of the material on the slope. Slopes are a product of tectonic uplift, and their strength is determined by the type of material on the slope and its water content. Rock strength varies widely and is determined by internal planes of weakness and their orientation with respect to the slope. In general, the more water contained by the slope material, the greater the likelihood of failure. This is especially true for unconsolidated sediments, where excess water pushes against the grains. Addition of water is the most common trigger of mass wasting and can come from storms or rapid snow melt.

15.2 Classification of Mass Wasting

The key criteria for classifying mass wasting are the nature of the movement that takes place, the type of material, and the speed of the material movement. Mass wasting events can be a precipitous fall of rock through the air, material sliding as a solid mass along either a plane or a curved surface, or internal flow of material as a viscous fluid. The type of material influences the mass movement, specifically whether it is solid rock or unconsolidated sediments. Slope failures can have translational

(planar) or rotational (curved) rupture surfaces. The important types of mass wasting are creep, slump, slide, fall, and debris flow or mudflow.

15.3 Preventing, Delaying, and Mitigating Mass Wasting

We cannot prevent mass wasting, but we can delay it through efforts to strengthen the materials on slopes. Strategies include adding mechanical devices such as rock bolts or ensuring that water in the slope materials can easily drain away. Such measures are never permanent but may be effective for decades or even centuries. We can also avoid practices that make matters worse, such as cutting into steep slopes or impeding proper drainage. In some situations, the best approach is to mitigate the risks associated with mass wasting by constructing shelters or diversionary channels. In other cases, where slope failure is inevitable, we should simply avoid building in that location.

Review Questions

1. In the scenario shown here, the gravitational force on the unconsolidated sediment overlying the point marked with an X is depicted by the black arrow. The red arrow in the diagram depicts the shear strength of the sediment.

Source: Steven Earle (2015) CC BY 4.0
View source
https://opentextbc.ca/geology/

   a. Draw in the two arrows that show how this force can be resolved into the shear force (along the slope) and the normal force (perpendicular to the slope).
   b. Assuming that the relative lengths of the shear force arrow (which you drew in question 1), and the shear strength arrow are indicative of the likelihood of failure, predict whether this material is likely to fail or not.
   c. After several days of steady rain, the sediment becomes saturated with water and its shear strength is reduced by 25%. What are the likely implications for the stability of this slope?
   d. Did you consider the affect of the additional weight of the water on the gravitational force acting on the slope in your answer to (c)? Does this change your answer?

2. In the diagrams shown here, a road cut is constructed in sedimentary rock with well-developed bedding. On the left, draw in the orientation of the bedding that would represent the greatest likelihood of slope failure. On the right, show the orientation that would represent the least likelihood of slope failure.

` Source: Steven Earle (2015) CC BY 4.0 View source https://opentextbc.ca/geology/

3. Explain why moist sand is typically stronger than either dry sand or saturated sand.

4. In the context of mass wasting, how does a flow differ from a slide?

5. If a large rock slide starts moving at a rate of several metres per second, what is likely to happen to the rock, and what would the resulting failure be called?

6. In what ways does a debris flow differ from a mudflow?

7. In the situation described in the chapter regarding lahar warnings at Mt. Rainier, the residents of the affected regions have to assume some responsibility and take precautions for their own safety. What sort of preparation should the residents make to ensure that they can respond appropriately when they hear lahar warnings? What other considerations do officials have to make in their emergency plan, other than just sounding a warning?

8. What factors are likely to be important when considering the construction of a house near the crest of a slope that is underlain by glacial sediments?

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-NC-SA 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 16. Earth-System Change

Adapted by Karla Panchuk from *Physical Geology* by Steven Earle

Figure 16.1 Antarctic Peninsula. Antarctica was not always covered by ice. A change in the Earth system triggered the onset of Antarctic glaciation approximately 40 million years ago. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Photo: Liam Quinn (2011). See Appendix C for more attributions.*

Learning Objectives

After reading this chapter and answering the review questions at the end, you should be able to:

- Explain why it is useful to think of Earth as a system.
- Describe the effect of feedbacks on the climate system.
- Explain the difference between weather and climate.
- Describe the climate-forcing mechanisms related to insolation, heat transport over Earth's surface, and changes in the atmosphere's energy budget.
- Explain the difference between direct and proxy data about Earth's climate, and give examples of each.
- Describe the role of computer models for understanding the Earth system.
- Summarize the history of human influence on the Earth system.
- Explain how carbon isotopes link rising atmospheric CO_2 levels to fossil fuels burned by humans.
- Describe how humans have affected the present-day carbon cycle and why the human contribution is significant.
- Summarize the kinds of observations that show signals of present-day climate change.
- Describe the current and projected state of the Earth system in the Anthropocene Epoch.

The Only Constant Is Change

If one thing has been constant about the Earth system over geological time, it is unceasing change. In the geological record of climate, sedimentary deposits provide evidence of glaciations in the distant past; the chemical characteristics of sea-floor sediments tell about periods of extreme warmth. The Earth-system, and thus Earth's climate, has not only changed frequently, but also with large temperature fluctuations. Today's mean global temperature is approximately 16°C. During Snowball Earth episodes more than 600 million years ago, when Earth's surface was frozen from pole to pole (or nearly so), the global mean was as cold as -50°C. At various other times in Earth history, it has been close to 30°C.

Part of this chapter addresses natural processes of climate change, how they work, and how we know what Earth's past climate was like. Geologists study those natural climate-change processes to understand how human-caused, or **anthropogenic**, changes to the Earth system might affect the climate in the future, and how much the climate has changed over the time that humans increased their influence on the Earth-system. The rest of the chapter addresses what has been learned by asking those questions.

16.1 What is the Earth System?

Earth can be characterized in terms of its "spheres." The **atmosphere** is the envelope of gas surrounding the planet. The **hydrosphere** is the water on the planet, whether in oceans, rivers, glaciers, or the ground. The **biosphere** comprises living organisms. The **lithosphere** is the rocky outer shell of the planet.

Components of these spheres interact constantly, with processes occurring in one sphere having an impact in other spheres. Cycles such as the water cycle or the carbon cycle constantly move matter and energy between spheres. Taking an **Earth-system** approach—looking at how the spheres are connected—is a way to account for the web of interactions responsible for the "big picture" of the Earth that we know.

The climate change related to the opening of the Drake Passage (Figure 16.2) is a good example of why a system of interactions is needed to understand how Earth works. The Drake Passage (bottom left map) is the gap between the southern-most tip of South America, and Antarctica.

Prior to 40 million years ago, the Drake Passage did not exist (top left map), and neither did the Antarctic ice cap. The arm of land connecting South America and Antarctica allowed warm ocean currents (red arrows) to carry heat from the equator to Antarctica. When the gap opened up, a new cold-water current formed (blue arrows) that blocked warm water from reaching Antarctica. Without the warm current, Antarctica froze over.

There were many interconnecting processes within the Earth-system (Figure 16.2, right) that drove glaciation in Antarctica. First, heat energy within Earth drove plate tectonics (lithosphere), making it possible for South America to separate from Antarctica. This impacted ocean currents (hydrosphere), and ultimately how water was stored on Antarctica (hydrosphere) by changing the climate of Antarctica.

In the Earth system, nothing happens in isolation. The change in the climate of Antarctica had a global impact. The ice cap on Antarctica increased Earth's **albedo**, the reflectiveness of Earth's surface. The more reflective Earth's surface, the more of the sun's light is reflected back into space without heating Earth. This caused even more ice to form, and cooled the planet as a whole.

When Earth cools, the change in temperature has a cascade of effects including changing precipitation patterns (hydrosphere), and changing the characteristics of habitats (biosphere). When habitats cool, organisms needing more warmth will migrate closer to the equator. This is true of plants as well as other forms of life.

Ice is not the only type of land cover that affects albedo—forests do as well. Forests also increase local atmospheric moisture levels through transpiration, when they release water vapour into the atmosphere. Local temperature and moisture differences also affect rainfall patterns on top of larger-scale changes resulting from cooling.

Figure 16.2 An example of Earth-system interactions. The opening of the Drake Passage changed how ocean currents moved heat around the planet, and cooled Antarctica until it froze over. *Source: Karla Panchuk (2018) CC BY-NC-SA 4.0. Map: Modified after Woods Hole Oceanographic Institution. See Appendix C for more attributions and terms of use.*

The chain of events in summarized in Figure 16.2 is only a broad overview of all of the consequences of opening a gap between South America and Antarctica. For example, it does not include the effects of what a change in the types of plants in a location does to local weathering and erosion. Trees can accelerate weathering, releasing more nutrients from rocks into runoff, which can affect algae blooms in water bodies, which in turn reduces oxygen levels in the water, which affects organisms living in the water that rely on oxygen.

Trees growing along a river can also slow the rate of erosion, reducing the amount of sediment in the river, and ultimately the rate of development of a delta at the river's mouth. Deltas undergo subsidence as accumulated sediments are compressed, so if the sediment supply is reduced, parts of the delta may become flooded, changing the extent of wetlands. Wetlands with waters depleted in oxygen can prevent plant material from decaying and releasing their carbon back into the carbon cycle as carbon dioxide. Changing atmospheric carbon dioxide levels alters the way energy moves through Earth's atmosphere, and affects Earth's surface temperatures.

The short version of why it's important to look at Earth as a system is that everything is connected, so that a change in one part of the system can ripple through the rest of the system and have effects well beyond any one location or time.

Feedbacks Amplify or Diminish Earth-System Change

The web of interactions in the Earth system is complex, but there is yet another level of complication. Sometimes a change in the Earth system can trigger other changes that have the effect of amplifying the original change, or diminishing it. The series of interactions that amplify or diminish a change are called **feedbacks**. A feedback that amplifies change is called a **positive feedback**. A feedback that diminishes the size of a change is called a **negative feedback**.

In the events related to the glaciation of Antarctica, the formation of ice is an example of a positive feedback. Ice formation was caused by cooling, but it triggered even more cooling by reflecting sunlight away from Earth's surface. This is called **ice albedo feedback**. An example of a negative feedback is plant growth. Plants need CO_2 to make food, so as long as the plants have enough nutrients and water, and temperatures are still suitable, increasing CO_2 in the atmosphere could increase plant growth. Plant growth would draw down atmospheric CO_2, so that there would be less warming than would otherwise be expected from the initial rise in atmospheric CO_2 levels.

Misconceptions About Feedbacks

There are two common misconceptions about feedbacks. One misconception is that positive feedbacks result in changes that are good, and negative feedbacks result in changes that are bad. In fact, whether a feedback is positive or negative is unrelated to whether or not the change would be considered a good thing. For example, if a feedback accelerates warming and makes an ecosystems uninhabitable for animals that used to live there, it would still be a positive feedback even though it had a negative impact on the animals in that ecosystem. A feedback that slowed the rate of warming and gave the animals time to adapt would still be considered a negative feedback even though it helped the animals to survive.

Another misconception is that a positive feedback always results in some value increasing (e.g., a rise in temperature), and a negative feedback results in a decrease in that value. Positive feedbacks can cause a value to decrease (e.g., as ice forms more sunlight is reflected, leading to decreased temperatures), and negative feedbacks can cause a value to increase. What matters is whether the initial change is amplified or reduced, not which way the numbers are changing.

Feedbacks and Instability in the Earth System

The potential for sudden extreme changes in the Earth system depends on what feedbacks are available. At times when Earth's climate was much warmer than today, no glaciers were present. When the climate is much cooler, a relatively small decrease in temperature could be enough to start the formation of ice and trigger the ice albedo feedback. However, if the climate is much warmer, the same decrease in temperature would not cool Earth enough to trigger the ice albedo feedback, and further climate cooling would be avoided. The reverse is also true- if warming occurs in a climate that is cold enough for glaciers to form, some of that ice might melt, reducing the albedo of Earth's surface, and permitting even more warming. On the other hand, if the climate is already too warm for ice to exist, a small amount of warming won't be amplified in the same way.

The albedo effect is not the only feedback that can make cooler climates less stable. Melting of **permafrost** (sediment that remains frozen year round) can also have an impact. Frozen soil contains trapped organic matter that is converted by micro-organisms to CO_2 and methane (CH_4) when the soil thaws. Both these gases contribute to warming when they accumulate in the atmosphere. Additional warming can cause even more permafrost to melt, permitting even more activity by micro-organisms, and releasing more CO_2 and CH_4.

Either of these feedbacks is enough on their own to accelerate climate change, but when they are both present together, the effect is even stronger. What this means is that the conditions in the Earth system *before* a change happens—called the **initial conditions**—play an important role in determining the impact of any changes that occur. A change that would have little impact under one set of initial conditions could have far reaching effects under another. Thinking of Earth as a system is a way to factor in the initial conditions. Otherwise we would be very puzzled why a small rise in global temperatures at one time in Earth history could have almost no discernible effect, but the same rise in temperatures at another time could lead to profound change.

16.2 Causes of Climate Change

What Is Climate?

Our day-to-day experience of the Earth system is in the form of the conditions we experience at Earth's surface. The daily conditions that we think of as **weather**—the temperature, presence or absence of precipitation, winds, humidity, and so on—are a snapshot of the state of the Earth system at a particular

instant in time and in a particular location. The weather that we get is variable, but in Saskatchewan most people would not be surprised to experience summer days with temperatures of 20 °C to 30 °C, and winter days with temperatures between –20 °C to –30 °C. Our notion of what summers and winters are generally like reflects our understanding of Saskatchewan's **climate**. If we get a day in July with a daytime high of 10 °C, that would seem like unusually cold *weather* because we know it is uncharacteristic of the *climate* over all.

We characterize the climate by collecting data about the weather every day, and then calculating the average conditions over a period of decades. The Government of Canada provides averages for the periods 1961 to 1990, 1971 to 2000, and 1981 to 2010 in an online database that is searchable by geographic location or station. (See the *Resources* section for the link.) Data measured at Saskatoon's Diefenbaker International Airport show that the average annual temperature from 1981 to 2010 is 0.6 °C higher than the annual average from 1961 to 1990, due warmer conditions in the winter and early spring (Figure 16.3).

Figure 16.3 Average temperatures for the periods 1961 to 1990, and 1981 to 2010 measured at Saskatoon's Diefenbaker International Airport (YXE). *Source: Karla Panchuk (2018) CC BY 4.0. Data from Canadian Climate Normals (see Resources).*

The *climate* as represented by the 1961 to 1990 interval was slightly cooler than the *climate* represented by the 1981 to 2010 interval. People who lived in Saskatoon between 1961 and 2010 may or may not have a sense that the *weather* they experienced from day to day was different for those intervals. In fact, some may have the record high of 35.3 °C on September 4, 1978 seared into their memory, and feel that Septembers just aren't as hot as they used to be. They would be correct that as of 2017, there are no September temperatures recorded at the Diefenbaker International Airport weather station with a daytime high greater than 35.3 °C. But if that gave them the impression that Septembers are cooler *on average* today than in the past, that would not be consistent with the data.

Climate-Forcing Mechanisms

A **climate-forcing mechanism** is a process that causes climate to change. Climate forcings work by initiating changes in how heat energy moves into, through, and out of the Earth system. When we discuss a particular climate change event, the climate-forcing mechanism is what initiated the change. Feedbacks also alter climate, but we want to know what triggered the feedbacks in the first place.

Climate Forcing by Changes in Insolation

Insolation, or **incoming solar radiation**, refers to how much of the sun's energy reaches Earth's surface in a given period of time. Insolation is measured in Watts per square meter (W/m^2).

Long-term Solar Evolution

Over the long term (billions of years), stars like our sun become larger, brighter, and hotter (Figure 16.4). Earth receives 40% more heat from the sun today than it did 4.5 billion years ago. In Figure 16.4, the blue

Now arrow shows the sun's current point in its life history. Although the blue arrow appears to indicate an instant in time, the time interval reflecting the duration of human existence on Earth is but a tiny fraction of the width of the line. As far as human experience is concerned, the long-term evolution of the sun is so slow that it has made no difference at all on insolation for the entire time humans have existed.

Figure 16.4 The life history of a star, from condensation of a nebula, to expansion to a red giant, and ending as a white dwarf. *Source: Karla Panchuk (2017) CC BY 4.0. Modified after Oliver Beatson (2009) Public Domain*

Orbital Cycles

Insolation is also affected by cyclical changes in Earth's orbit and rotation. Over intervals of approximately 100,000 years, the **eccentricity** of Earth's orbit changes. Eccentricity is a measure of how elliptical a circle is. Higher eccentricity means that the orbit is more elliptical (Figure 16.5, left, blue orbit), whereas lower eccentricity means the orbit is more circular (Figure 16.5, left, red orbit). **Eccentricity** is important because when it is high, the Earth-sun distance varies more from season to season than it does when eccentricity is low.

Over intervals of approximately 41,000 years, the **obliquity** of Earth's axis of rotation changes (Figure 16.5, middle). This results in a nodding motion that alters how directly the sun shines on Earth's poles. When the angle is at its maximum (24.5°), Earth's seasonal differences are accentuated. When the angle is at its minimum (22.1°), seasonal differences are minimized.

Cycles of **precession** happen over intervals of approximately 20,000 years, causing Earth's axis of rotation to wobble (Figure 16.5, right). This means that although the North Pole is presently pointing to the star Polaris (the pole star), in 10,000 years it will point to the star Vega.

Figure 16.5 Cycles in Earth's orbit. Left: The shape of Earth's orbit (its eccentricity) changes over 100,000 year cycles from more circular to more elliptical. Middle: Over 41,000 year periods, Earth's axis of rotation nods toward and away from the sun. Right: Over 21,000 year cycles, Earth wobbles on its axis of rotation. *Source: Karla Panchuk (2017) CC BY 4.0. Modified after Steven Earle (2015) CC BY 4.0. See Appendix C for more attributions.*

The importance of eccentricity, tilt, and precession to Earth's climate cycles (now known as **Milanković Cycles**) was first pointed out by Yugoslavian engineer and mathematician Milutin Milanković in the early 1900s. Milanković recognized that although the variations in the orbital cycles did not affect the *total* amount of insolation that Earth received, it did affect *where* on Earth that energy was strongest. Glaciations are most sensitive to the insolation received at latitudes of approximately 65°. As continents are configured today, this is most significant at 65° N, because there is almost no land at 65° S.

The most important factors are whether the northern hemisphere is pointing toward or away from the sun at its closest or farthest approach, and how eccentric the sun's position is in Earth's orbit. For example, if the northern hemisphere is at it farthest distance from the sun during summer (Figure 16.6, top), this means cooler summers. If the northern hemisphere is at its closest distance to the sun during summer (Figure 16.6, bottom), this means hotter summers. Cool summers — as opposed to cold winters — are the key factor in the accumulation of glacial ice, so the upper scenario in Figure 16.6 is the one that promotes glaciation. This factor is greatest when eccentricity is high.

Figure 16.6 Effect of precession on insolation in the northern hemisphere summers. In (a) the northern hemisphere summer takes place at greatest Earth-sun distance, so summers are cooler. In (b) (10,000 years or one-half precession cycle later) the opposite is the case, so summers are hotter. The red dashed line represents Earth's path around the Sun. *Source: Steven Earle (2015) CC BY 4.0*

The effects of all three cycles are evident in geochemical climate data. Figure 16.7 shows the "signals" for obliquity (A), eccentricity (B), and precession (C) over a period from 800,000 years in the past, to 800,000 years in the future. The vertical black line running down the middle of the diagram marks the present day. When the insolation from all three signals is determined, the result is a more complex waveform (D) with times of low variation in insolation, and times with higher variation in insolation.

Figure 16.7 Comparison of orbital cycles, insolation, and climate data for a 1.6 million year period. *Source: Karla Panchuk (2017) CC BY-SA 4.0, modified after Incredio (2009) CC BY 3.0. See Appendix C for more attributions.*

The graphs E and F are climate information measured in microfossils dwelling at the ocean floor (E) and in water from ice cores (F). Peaks in temperature in F correspond to peaks in the oxygen isotope record in E, which indicate that it was warmer, there was less ice, and sea level was higher. Troughs are times when Earth was deep within an ice age. It was cooler, there was ice on land, and sea level was lower.

The vertical dashed lines on the left-hand side of Figure 16.7 mark the times of peak warm temperatures and allow for comparison of the timing of the temperature peaks with the timing of the orbital cycles. Peak temperature events are approximately 100,000 years apart, suggesting that the eccentricity cycle might be the most important contributor. Indeed, in B most (but not all) of the peak temperature events correspond to a time when Earth's orbit was at or near peak eccentricity for that cycle.

It is tempting to conclude that eccentricity is the most important orbital cycle for climate change over all. However, this pattern only began a little over 1 million years ago. For 1.5 million years before that, the 41,000-year obliquity cycle seems to dominate insolation cycles.

In general, times of warmest or coolest temperatures don't line up perfectly with orbital cycles. There is no one orbital cycle that is most important for all of Earth history. It is also the case that changes in insolation due to orbital cycles are not sufficient to cause temperatures to change as much as the geological record says they have; feedbacks must be factored in to explain the observed temperature changes.

Sunspot Cycles

Sunspots are dark patches that appear on the surface of the sun as a result of intense local disturbances in the sun's magnetic field (Figure 16.8, left). Loops of plasma (gas with electrical charge, Figure 16.8, right) follow along magnetic field lines from one sunspot to another.

Figure 16.8 Sunspots. Left: Photograph of sunspots with dots representing the size of Earth and Jupiter for scale. Right: Plasma loops viewed in x-ray wavelengths jumping from one sunspot to another on the sun's surface. *Source: Left- NASA/Solar Dynamics Observatory (2012) Public Domain. Right- NASA/Solar Dynamics Observatory (2015) Public Domain*

Sunspots appear dark because they are lower-temperature regions on the sun's surface. For that reason you might think that more sunspots means a reduction in insolation. In fact, just the opposite is true, because sunspots are a side-effect of increased solar activity. Peaks in the number of sunspots counted annually since approximately 1870 (Figure 16.9, blue), coincide with peaks in measurements of solar energy output from the same time period (Figure 16.9, pink).

Sunspot cycles happen over approximately 11 year intervals, and the changes in insolation that occur during these cycles are relatively small. In the end the effect of sunspot cycles on climate can be lost amidst other factors. In Figure 16.9 there is no clear relationship between the sunspot cycles and the global average temperatures (in grey) reported for the same period.

Figure 16.9 Sunspot cycles. Peaks in the number of sunspots (blue) occur approximately every 11 years, and these correspond to peaks in solar energy output (pink). The influence of sunspot cycles is too small to have a clear impact on global average temperatures (grey). *Source: Karla Panchuk (2017) CC BY-SA 4.0. Sunspot records modified after D. Bice (n.d.) CC BY-SA 3.0. Global average temperature modified after Met Office (2015) Contains public sector information licensed under the Open Government Licence v1.0.*

Be Aware of Graph Scales

Figure 16.9 shows three kinds of data: temperatures, sunspot numbers, and solar energy flow. Each of these data sets is a different type of information, so each needs its own vertical axis. The vertical axes are scaled so that the data fill the area of the graph as much as possible. Stretching the vertical scale to fit the full plotting area makes it easier to see how well the peaks and troughs in each record line up with each other. Unfortunately, this can also skew our impression of the data. For example, in the period from 1880 to 1920, all three records have a similar vertical distance from peak to trough. In other words, all three records have approximately the same size of wiggles. This does not mean that the change in insolation from sunspot cycles was big enough to cause all of the variation in the temperature record. From this graph alone, there is no way to tell how much the change in insolation due to sunspot cycles mattered to global temperatures during the period 1880 to 1920.

Climate Forcing by Changes in Heat Transport

The ocean transports large amounts of heat around the Earth through a conveyor-belt-like system of currents. The ocean has surface currents that are driven by wind, but it also has deeper currents that are not wind-driven. The deeper currents behave like stacked rivers because they are different temperatures and have different salt contents, and therefore different densities. The differences in density between these water masses are what drive circulation. Circulation that is driven by density is called thermohaline circulation; *thermo* refers to heat and *haline* refers to salt.

To see how this works, consider the warm and saline Gulf Stream current (Figure 16.10, top). It flows northward past Britain and Iceland into the Norwegian Sea, and cools as it moves north, becoming denser. Its high salinity contributes to its density, and it sinks, or **downwells**, deep beneath the surrounding water, forming the **North Atlantic Deep Water** (NADW) current that flows south. Meanwhile, at the southern extreme of the Atlantic, very cold water adjacent to Antarctica also sinks to the bottom to become the **Antarctic Bottom Water** (AABW) current. The AABW flows north, beneath the NADW.

Figure 16.10 A simplified north-south cross-section through the Atlantic Ocean basin showing the different current layers. *Source: Steven Earle (2015) CC BY 4.0*

The water that sinks in the areas of deep water formation in the Norwegian Sea and adjacent to Antarctica moves very slowly at depth. It eventually resurfaces, or **upwells**, in the Indian Ocean between Africa and India, and in the Pacific Ocean, north of the equator (Figure 16.11).

Figure 16.11 Global thermohaline circulation patterns. Red lines are surface currents, and blue lines are deep currents. *Source: NASA Earth Observatory (2008) Public Domain*

Some ocean currents move warm water from the equator toward the poles. As in the example of the Drake Passage, the path of warm currents can have a significant impact on the climate of a region, and potentially of the planet as a whole. Processes that disrupt the density of seawater can slow or stop currents, preventing warm water from reaching higher latitudes. The recovery from the last ice age is characterized by sudden returns to glacial conditions over as little as 3 years. This is thought to be the result of enormous glacial lakes forming on continents as the glaciers melted, then being suddenly released into the ocean by a burst ice dam. The glacial water would be very cold, but it would also be fresh, making it less dense than the ocean water. The fresh glacial water would form a cap and slow the downwelling conveyor belt at high latitudes.

Scientists are trying to determine the current and past state of the Atlantic-basin system of circulation, called the **Atlantic Meridional Overturning Circulation** (**AMOC**), to tell whether it is changing in response to warming and adding fresh water from melting ice sheets. The AMOC varies considerably on decadal cycles because of cycles in the wind patterns in the Atlantic, so it is important to distinguish these cyclical changes from any longer-term underlying changes.

Because of these studies, we have an idea of what the physical properties of the Atlantic Ocean look like when circulation is stronger or weaker. When circulation slows, the density is lower in the downwelling regions (Figure 16.12, blue patch in the Labrador Sea). The density is higher south of this region, along the eastern coast of the United States and southern Canada (Figure 16.12, orange patch). Model simulations are used to confirm that the changes in density we observe are consistent with how we understand the circulation system to work.

Measurements of the actual flow rate at depth in the Atlantic Ocean (Figure 16.12, right, blue dots) confirm that density decreases in downwelling regions (black and red lines) when circulation slows. A more recent

study (Caesar et al., 2018) has shown that observed and model temperatures follow the same pattern, with cooling where the blue patches are in Figure 16.12, and warming in the region of the orange patches. This is to be expected because the slowdown in circulation affects how heat is moved northward.

Figure 16.12 Using changing density to track circulation in the Atlantic Ocean. Left- Density calculated from measurements of temperature and salinity in a layer between 1,000 m and 2,500 m depth in the Labrador Sea (black box). Middle- Model results used to see what change in density can be expected. The model shows the same general relationship, with lower density in the Labrador Sea, and higher density to the south. Note that density units are in kg/m^2 rather than kg/m^3 because they are integrated over the layer. Right- Changing density in the Labrador Sea over time. Red and black lines show changes in density from two different data sets. In general, the peaks and troughs of these data sets match up. Blue dots are measurements of the rate of circulation from a project that began collecting data in 2004. See the References section for more information about the relevant studies. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Modified after Jon Robson (2013) CC BY-SA 4.0*

Is Atlantic Circulation Slowing Down More than Usual?

Changes in the Atlantic meridional overturning circulation (AMOC) happen from decade to decade. To know whether circulation is changing compared to what is normal, it is necessary to get information about what circulation looked like in the past. This is difficult to do, because measurements of circulation rates, temperature, and salinity don't go back as far as we need them to.

In a new study, Thornalley et al. (2018) have used geochemical analyses of microfossils to build a longer-term record of temperatures, then used that record to look for the temperature "fingerprint" of slowing AMOC, an increasing difference in the temperatures of surface waters compared to deeper waters in the downwelling zone (Figure 16.13, top). They observe a longer-term cooling trend beginning at the close of the Little Ice Age, suggesting that less heat is being moved toward the downwelling zone (labelled A on the globe in Figure 16.13).

Figure 16.13 Long-term record of Atlantic Meridional Overturning Circulation (AMOC). Top- Temperature fingerprint of circulation determined using geochemical analyses of marine microfossil shells. Results show the difference between temperatures measured at depth at A on the globe, and temperatures measured near the surface at B. Cooling at A relative to B is indicative of weakening AMOC. Bottom- Changes in silt grain-size used to show changes in the velocity of currents at depth at C on the globe. A smaller average grain size means a slower current, and weaker AMOC. *Source: Karla Panchuk (2018) CC BY 4.0. Modified after Thornalley et al. (2018). Locator globe modified after Reisio (n.d.) Public Domain*

They also measured the size of silt grains on the sea floor, above which a southward-moving component of the AMOC system flows (location labelled C on the globe in Figure 16.13). Grain size is used as a substitute for a direct measure of ocean current velocity because the velocity determines what grain size can be carried. They identified a lower average grain size over the past ~150 years compared to the average grain size from earlier (Figure 16.13, bottom, dashed lines). The authors of the study note that the average grain size changes more during cold events in the Northern Hemisphere (the Dark Ages Cold Period and

the Little Ice Age).

Thornalley et al (2018) conclude that the AMOC has been weaker on average during the past ~150 years than during the previous ~1,500 years. However, they cannot say for sure how much of that change is from melting that occurred at the close of the Little Ice Age, from melting triggered by warming since the Industrial Revolution, or some combination of the two. Direct measurements of density and current velocity tell us that as of 2017, the AMOC continues to weaken.

Plate Tectonics and Heat Transport

The opening of the Drake Passage is one example of how plate tectonic changes can affect ocean heat transport, and therefore climate. Plate tectonic changes that build or break up continents also play a role. When continents become large, ocean currents warm their margins, but the interiors can be much cooler. Anyone living on the Canadian prairies who has shivered through -40 °C temperatures in the winter, while watching news reports of rain in Vancouver will be familiar with this effect. When the supercontinent Gondwana was over the south pole approximately 300 million years ago (Figure 16.14), this triggered an ice age. The build-up of ice was hastened by the ice albedo feedback effect.

Figure 16.14 Glaciation on the supercontinent Gondwana. Paleogeographic reconstruction for 306 million years ago. *Source: C. R. Scotese, PALEOMAP Project (*www.scotese.com*). See Appendix C for terms of use.*

Short-Term Cycles in Heat Transport: El Niño Southern Oscillation

The **El Niño Southern Oscillation** (**ENSO**) operates on a much shorter timescale than climate forcings driven by plate tectonics or orbital cycles, alternating between **El Niño** and **La Niña** events on timescales of between two and seven years (Figure 16.15).

Figure 16.15 Variations in the ENSO index from 1950 to 2015. *Source: Steven Earle (2015) CC BY 4.0. Modified after Klaus Wolter/ NOAA (n.d.) Public Domain*

Under normal conditions, strong winds blowing westward across the Pacific cause water to pile up in the western Pacific. This forces deeper colder water to the surface in the eastern Pacific (Figure 16.16, left). During La Niña events, further intensification of winds causes even more cold water to upwell. During an

El Niño event, the winds weaken, allowing water to flow back to the east (Figure 16.16, right). The cold water settles deeper once again, meaning that warmer water is present along the eastern margin of the Pacific Ocean.

Figure 16.16 El Niño Southern Oscillation (ENSO) cycles are driven by changes in wind patterns that affect the distribution of warm and cold water in the Pacific Ocean. *Source: Karla Panchuk (2018) CC BY 4.0. Modified after Fred the Oyster and NOAA/PMEL/TAO Project Office.*

ENSO events affect weather on a global scale (Figures 16.17 and 16.18). In western Canada, El Niño years have warmer than average winters, whereas La Niña years have cooler than average winters.

Figure 16.17 El Niño climate impacts. *Source: NOAA Climate.gov (n.d.) Public Domain*

Figure 16.18 La Niña climate impacts. *Source: NOAA Climate.gov (n.d.) Public Domain*

Climate Forcing by Changes in the Atmosphere's Energy Budget

Earth's atmosphere regulates climate by controlling how much energy from Earth's surface escapes to space, and how much of the sun's energy reaches Earth's surface.

Albedo

Albedo is a measure of the reflectivity of a surface. Earth's various surfaces have widely differing albedos, expressed as the percentage of light that reflects off a given material. This is important because most solar energy that hits a very reflective surface is not absorbed and therefore does little to warm Earth. Water in the oceans or on a lake is one of the darkest surfaces, reflecting less than 10% of the incident light. Clouds and snow or ice are among the brightest surfaces, reflecting 70% to 90% of the incident light (Figure 16.19).

Figure 16.19 Typical albedo values for Earth surfaces. Surfaces with low values reflect less light than surfaces with high values. *Source: Steven Earle (2015) CC BY 4.0*

Albedo, Feedbacks, and the Acceptance of Milanković Cycles as a Climate Forcing Mechanism

When Milanković published his hypothesis in 1924, it was widely ignored, partly because it was evident to climate scientists that the forcing produced by the orbital variations alone was not strong enough to drive the climate changes of the glacial cycles. Those scientists did not recognize the power of positive feedbacks. It wasn't until 1973, 15 years after Milanković's death, that sufficiently high-resolution data were available to show that the Pleistocene glaciations were indeed driven by the orbital cycles, and it became evident that the orbital cycles were just the first step, initiating a range of feedback mechanisms that made the climate change, many of which were related to albedo.

Consider the following:

- When large volumes of ice melt — such as the continental ice sheets of Antarctica and Greenland, as well as alpine glaciers— this decreases albedo. More solar energy is then absorbed by land, amplifying the increase in temperature.
- When sea ice melts and exposes water, the albedo of the exposed area decreases drastically, from approximately 80% to less than 10%. Far more solar energy can be absorbed by the water compared to the previous ice cover, amplifying the temperature increase.
- Sea level rises when ice and snow melt on land, and because seawater expands when heated. Higher sea level means a larger proportion of the planet is covered with water, which has a lower albedo than land. More heat is absorbed, amplifying the temperature increase. Since the last glaciation, a rise in sea level of approximately 125 m has flooded vast areas of land.

Exercise: Albedo Impacts of Vegetation Changes

Changes in climate can cause forests to be replaced by grasslands, which have higher albedo than dark forest cover. If deserts expand, vegetated areas can be replaced by higher-albedo sand. Many human activities affect albedo, including adding urban surfaces to an environment, and planting crops. Figure 16.20 shows a forest that has been clear-cut. If a clear-cut has an albedo similar to that of sand, how would clear cutting change the albedo of the area?

Figure 16.20 A clear-cut near Eugene, Oregon. *Source: Calibas (2011) CC BY-SA 3.0*

Note that trees cool their environment through transpiration, when they release water vapour from their leaves. Changes in local temperatures when trees are clear-cut also include the effects of reduced evaporative cooling. Changes in vegetative cover also affect the rates of CO_2 uptake by plants.

Greenhouse Gases (GHGs)

All molecules vibrate at various frequencies and in various ways, and some of those vibrations take place at frequencies within the range of the infrared radiation that is emitted by Earth's surface. Gases with two atoms, such as O_2, can only vibrate by stretching (back and forth; Figure 16.21 left), and those vibrations are much faster than that of IR radiation. Gases with three or more atoms (such as CO_2) can vibrate in other ways, such as by bending (Figure 16.21 right). Those vibrations are slower and allow the molecules to absorb and release infrared radiation.

Figure 16.21 Molecules with two atoms (top) vibrate differently from molecules with more than two (bottom), and this determines whether a gas will be a greenhouse gas or not. *Source: Steven Earle (2016) CC BY 4.0*

When infrared radiation interacts with CO_2 or with one of the other GHGs, the molecular vibrations are enhanced because there is a match between the wavelength of the light and the vibrational frequency of the molecule. This makes the molecule vibrate more vigorously, heating the surrounding air in the process. These molecules also emit infrared radiation in all directions, some of which reaches Earth's surface. Heating due to the vibrations of greenhouse gas molecules is called the **greenhouse effect**. Water molecules (H_2O), and methane molecules (CH_4) also interact with infrared radiation when they vibrate, so they are greenhouse gases as well.

Ice core records show that over the last 800,000 years, rapid cycles into and out of glacial temperatures are associated with similarly-timed cycles in atmospheric CO_2 levels (Figure 16.21).

Figure 16.22 Variations in atmospheric CO_2 levels and temperature over the last 800,000 years. Top- CO_2 concentration from ice core data (Lüthi et al., 2008). Dashed line: recent atmospheric CO_2 from the Mauna Loa Observatory. Bottom: Temperature from oxygen isotope measurements of water in ice cores (Jouzel et al., 2008). Upper dashed line: 2016 global surface temperature from NASA/GISS. Lower dashed line: average temperature for the past 1000 years. *Source: Karla Panchuk (2018) CC BY 4.0, modified after National Research Council (2010). Click the image for terms of use.*

Earth-System Response Time

You might have noticed that for most of the 800,000-year record in Figure 16.22, there is a fairly consistent relationship between scale of change in atmospheric CO_2 levels and the resulting change in temperature. You might also have noticed that compared to most of the record, the rise in temperature since 1950 is unexpectedly small, given the increase in atmospheric CO_2 levels since that time. The reason for the relatively small temperature increase in response to the recent CO_2 increase is in large part because the recent rise in CO_2 is happening far more rapidly than other parts of the Earth system can respond. The ocean in particular is slowing down the response.

The ocean takes up heat from the atmosphere, and thus helps to determine surface temperatures. A relatively cool ocean can take up more heat from the atmosphere, reducing warming. The fastest way for the ocean as a whole to take up heat is through the "stirring" that happens with ocean circulation, but circulation happens on thousand-year timescales. The slow rate of circulation means that centuries from now, there will still be cool water rising up from the deep ocean that has yet to be exposed to the warmer surface conditions. At other times in the 800,000-year record, changes in CO_2 levels happened on timescales much closer to those at which the ocean takes up heat.

Atmospheric Effects of Volcanic Eruptions

Volcanic eruptions don't just involve lava flows and exploding rock fragments. Eruptions also release particles and gases into the atmosphere. Important volcanic gases include water vapour, CO_2, and sulphur dioxide (SO_2). Volcanic CO_2 emissions can contribute to climate warming if a greater-than-average level of volcanism is sustained over a long time. At the end of the Permian Period, the massive Siberian Traps were produced by eruptions lasting at least a million years. Large quantities of CO_2 were released, warming the climate and triggering a cascade of Earth-system responses. The end of the Permian Period at 252 Ma is marked by the greatest mass extinction in Earth history.

Over the shorter term, however, volcanic eruptions can have the opposite effect, cooling the climate. SO_2 reacts with water in the atmosphere to make droplets of sulphuric acid. The sulphuric acid droplets scatter sunlight, reducing how much of the sun's energy can reach Earth's surface. They also affect cloud formation. The volcanic cooling effect is relatively short-lived, because the particles settle out of the atmosphere within a few years.

Exercise: Climate Change at the End of the Cretaceous Period

The large extraterrestrial impact at the end of the Cretaceous Period 66 Ma ago is thought to have produced a massive amount of dust, which may have remained in the atmosphere for several years. It may also have produced a great deal of CO_2. What do you think would have been the short-term and longer-term climate-forcing implications of these two factors?

constructed on limestone supported by beds of gypsum and anhydrite. Gypsum and anhydrite are soluble in water, and the gypsum and anhydrite beneath the dam are rapidly dissolving away. This was the case prior to construction of the dam. However, once the dam was filled, the increased water pressure began to force water through the formations much faster, accelerating dissolution. Ongoing measures to fill gaps with grout are required, or else there is a grave risk of catastrophic failure, placing nearly 1.5 million people at risk.

16.3 Methods for Studying Past Climate

Whereas *weather* refers to day-to-day variations in temperature, precipitation, winds, and so on, *climate* refers to long-term trends in weather patterns (over decades or more). The term **paleoclimate** refers to Earth's climate in the past. The information we have about Earth's past climates can be classified as **direct data** or **proxy data**. Direct data are information derived from first-hand observations of climate. Direct

data can be **instrumental data**, derived from tools designed to quantify observations, or from qualitative descriptions.

Proxy data are information derived from natural materials with characteristics that are affected by climate in a systematic way. This could also be said of some instrumental data: an alcohol thermometer uses the fact that the volume of alcohol changes in a consistent way in response to temperature. Proxy data rely on relationships that are also as systematic and consistent, but there are important differences:

- Instruments are designed so that the response of the instrument reflects only one characteristic of the environment (e.g., an alcohol thermometer is sealed so that only temperature affects the level of alcohol, not air pressure or evaporation), but proxy data may need to be carefully analyzed to account for other processes.
- Data from instruments are lost if observations are not made and recorded. A thermometer is constantly changing in response to temperature, and will not stay in a particular state once conditions change. Proxy records capture information about the environment in a way that can persist over the long term, for as long as the materials last in usable form. Some proxy records preserve information from billions of years ago.
- Instruments transform characteristics of the environment into information that we can access immediately (e.g., you can read a temperature directly from a thermometer), but materials from which we derive proxy data require processing, often with specialized laboratory equipment, to get data we can use.
- Instruments measure only what we design them to measure, and measure only as well as we design them to do so. Because proxy data come from materials that persist through time, it is possible to improve the techniques used to analyze materials, and redo earlier measurements. Discovering new proxies is an ongoing area of research, and regularly reveals ways to determine characteristics of past states of the Earth system that we didn't think were knowable.
- Instrumental records and other direct observations can be well constrained in time. In other words, we often know the exact dates the observations were made. Proxy data may come from materials that are difficult to pin down to an exact date, and additional information may be required to determine when the records were formed. Sometimes proxy data can only be extracted as an average over a number of years. The relevant question becomes whether the resolution of the record (how detailed it is) is high enough for the time frame of interest. For example, an annual average would be no use at all to understand monthly climate variations, but would provide exceptionally high resolution for geological processes operating over tens of thousands of years.

Types of Direct Data

Instrumental records of climate are those derived from tools such as thermometers, rain gauges, or satellite measurements of the extent of ice sheets. Instrumental records are a recent development, as the history of the Earth system goes. The oldest known temperature measurements cover the period from 1654 to 1670, and were made by monks and Jesuit priests who operated stations within a meteorological network supported by the Medici family of Florence.

Non-instrumental historical records of climate also exist, and cover periods of human history prior to the development of the climate-measuring tools we have now. With detective work, these can be used to paint a detailed picture of past climates. Non-instrumental historical records include written records about how long ice and snow were present in a particular year, when harvests occurred, when floods happened, and shipping records that report the extent of sea ice. Paintings of alpine glaciers give information about how far the ice extended, and this can be used to reconstruct temperatures.

The Challenges of Getting Climate Information from Historical Records

In their paper *Historical Climate Records in China and Reconstruction of Past Climates*, Jiacheng Zhang and Thomas Crowley used official Chinese records extending as far back as 1000 CE to get a detailed picture of climate. This involved transforming descriptions of weather events into a systematic scale. One challenge is defining the scale, but another is deciding what individual accounts actually mean. The authors point out that records of rain or drought can reflect the perceptions and generalizations of the people who wrote about the weather, rather than what actually happened. Consider the following description of rainfall in 1644 from *Diary of Qi Zongmin*:

"in Wyzhong there is no rain for six months from May"

Does this mean there was no rain at all prior to May, or just very little? Was it actually six months since there had been rain, or is that an approximation? How do we reliably translate different systems of time measurement into durations and dates, like "six months" or "May?" How can we tell for sure where a location is if place names or political boundaries change?

As the authors caution, "a great deal of cross-checking [is required] in order to arrive at a useful descriptive account of climate anomalies."

Sources of Proxy Data

Tree Rings

The study of tree growth rings for the purpose of understanding past states of the Earth system is called **dendroclimatology**. Temperatures and a history of drought or wet periods can be reconstructed from the widths of tree rings. Because tree rings form annually, these records can also be well constrained in time. The widths of tree rings reflect how fast the tree grows in a given season. There are factors other than temperature and moisture that affect growth rate, so a sampling strategy must be carefully designed to ensure confidence in climate reconstructions.

Stable Isotopes

Atoms have a nucleus made of protons and neutrons. The number of protons in the nucleus determines what element the atom is, and will always be the same for a given element. In contrast, the number of neutrons can vary for an element. Versions an element having different numbers of neutrons are the **isotopes** of that element.

Sometimes an atom has a number of neutrons that makes it unstable. Those atoms eventually break apart, releasing energy, and are called **radioactive isotopes**. The decay rate of radioactive isotopes is known, making it possible to use them to find the ages of natural materials. For example, carbon-14 dating makes use of the radioactive isotope of carbon, ^{14}C, which has eight neutrons instead of the usual number, six (the 14 refers to 6 protons + 8 neutrons).

For investigating Earth's past climate, **stable isotopes**, which do not decay, are used instead. Stable isotopes of the same element are measured in natural materials, and their ratios compared. Both isotopes are involved in the same chemical reactions and physical processes, but the slight difference in mass caused by one or two extra neutrons means that those processes are more likely to take up the lighter isotope than the heavier one. Some processes do this in such a particular way that evidence of their occurrence is left behind as a distinctive fingerprint in the stable isotope composition of materials formed in their environment.

The pair of isotopes used to reconstruct past temperatures are the oxygen isotopes ^{16}O and ^{18}O. The ratio of ^{18}O to ^{16}O in water is reflected in the calcium carbonate of shells that form in the water. The shells may

remain in the geologic record long after the water is gone, making it possible to know the oxygen isotope compositions, and thus temperatures, of water bodies that existed in the distant past.

Ice Cores

The ice in polar glaciers and mountain glaciers preserves a detailed snapshot of Earth's atmosphere and climate. A sample of Earth's atmosphere, including gases and particles, is captured and held within the ice, and buried beneath subsequent ice layers. The annual layers in the ice can be used to determine a timescale for the data. The gases in air bubbles trapped within ice (Figure 16.23) are analyzed to determine the chemical composition of the atmosphere at the time the gasses were trapped.

Figure 16.23 A researcher holds a fragment of ice from Antarctica. The dots in the fragment are air bubbles containing samples of Earth's past atmosphere. *Source: Atmospheric Research, CSIRO (2000) CC BY 3.0*

Ice cores (Figure 16.24) are cylinders of ice retrieved using a specialized drill bit. The cores are carefully packaged and stored in specially designed facilities (Figure 16.25) until they are analyzed.

Figure 16.24 A scientist weighs and measures a cylinder of core from the West Antarctic Ice Sheet before she packages it for transport. *Source: NASA/Lora Koenig (2010) CC BY 2.0*

Figure 16.25 National Science Foundation Ice Core Facility in Lakewood, Colorado. Cores are housed in tubes 1 m long. The main storage facility is kept at -36 ºC. Fortunately, scientists can examine the cores under much warmer conditions in a nearby room maintained at -24 ºC. *Source: U. S. Geological Survey/ Eric Cravens (n.d.) Public Domain*

Rock and Fossil Distributions

Earth can be divided into six main climate zones (Figure 16.26). The zones run roughly along lines of latitude, so that the climate zone changes as you move north or south of the equator. When the climate warms, the zones shift away from the equator; an area now in the boreal climate zone might have been in the warm temperate climate zone when Earth's climate was warmer. When the climate cools, the zones shift toward the equator.

Figure 16.26 Six climate zones of the Köppen-Geiger classification. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Modified after LordToran (2007) CC BY-SA 3.0*

Some rock types are characteristic of particular climate zones. For example, coal deposits are characteristic of a subtropical climate. Limestone with coral reef fossils is characteristic of a tropical climate. If a rock type is found outside of its climate zone, that might indicate a change in climate. Coal can be found near Estevan, Saskatchewan, now in the warm temperate climate zone. This suggests that at one time, a warmer climate resulted in a northward shift of the subtropical climate zone. Some of the oil in western Canada is present in pore spaces within ancient coral reefs. The warm temperate climate zone cannot have been at its present location when those reefs formed.

Fossils can be used similarly. If an organism lives in a habitat with a particular climate, then evidence that the organism has migrated away from the equator could indicate warming. Migration toward the equator could indicate cooling.

The study of pollen and plant spores, called **palynology**, is very helpful for determining the distribution of plants when evidence of larger plant parts (e.g., fossil leaves and bark) is absent. Pollen and spores are very tough, and will survive in the environment when other plant materials do not. A detailed record of pollen and spores, and hence of the climate zones in a particular location, can be derived from lake sediments. Lakes in climates with strong seasonality (a distinct difference in temperature as seasons progress) can accumulate distinct annual sediment layers, called **varves** (Figure 16.27). Each year is represented by a light layer and a dark layer. The light layers consist of sand and silt from spring runoff. The darker layers include organic matter accumulated during the year.

Figure 16.27 Varves in a core from Canoe Brook, Drummerston, Vermont. Each pair of light and dark layers represents one year. The top of the core is to the right. *Source: Karla Panchuk (2017) CC BY-NC-SA 4.0 (labels added). Modified after Jack Ridge/ North American Glacial Varve Project (2008). See Appendix C for terms of use.*

Varves can be counted to determine the age of the sediment, and the pollen and spores within the sediment can be extracted to see what types of vegetation were present at different times.

16.4 Computer Models of the Earth System

Earth-system interactions are so complex that it is next to impossible to follow all of the connections and implications without help. So, scientists use **Earth-system computer models** to assist. Earth-system models incorporate knowledge of the many components of the Earth system in a way that makes it possible to test how important any one change is.

Earth-system models vary in how many aspects of the Earth system they include, and how detailed their representations of those aspects are. Models are designed to answer particular kinds of questions so their performance can be optimized; a study that is concerned only with large-scale global changes might not require a model with a highly detailed representation of Earth's coastlines. It takes more time to run a complicated model, so this saves on computing resources.

When computer models are discussed, we acknowledge that there is a difference between measurements of the real world, and the output from the model. Modellers are careful to refer to measurements of the real word as *data*, and output from the model as *results*. This also helps to avoid confusion when comparing models to real-world measurements to gauge how realistic the model output is.

What Are Computer Models, Exactly?

Computer models describe natural phenomena using mathematical equations. On the most basic level, computer models take some quantity—whether heat, water, or the concentration of a pollutant—and calculate how it moves through a system. Sometimes they look only at how that quantity changes through time. A computer model of the water volume in a bathtub could be limited to looking at how rapidly water flows in through the tap, and how rapidly it flows out through the drain. But sometimes models look at how a quantity changes in space as well as through time.

A study of wind-driven currents in a lake must include information about the shape and depth of the lake to capture how friction at the lake bottom and along the sides affects water flow (Figure 16.28). Data about the lake shape and depth (Figure 16.28, top) is translated to a model grid (Figure 16.28, bottom). Calculations are done to see how wind and friction control how water moves into and out of each cell in the grid.

Figure 16.28 Set-up for a model of wind-driven current flow in Lake Ontario. Top: Map of Lake Ontario showing water depth and the location of current meters. Bottom: Grid used to translate water depth information for model calculations. *Source: Karla Panchuk (2002) CC BY 4.0. Based on the exercise described in Chapter 10 of Slingerland & Kump (2011).*

The model produces information about wave height (Figure 16.29, left) and shows the direction and speed of water flow across the lake using arrows of different sizes (Figure 16.29, right). If scientists are interested in how a pollutant would move around the lake, they can include the location where the pollutant is added, and how rapidly it is added, and track how it moves.

Figure 16.29 Height of the water surface (left) and current velocity (right) from a model of wind-driven flow in Lake Ontario. The length of current arrows shows the speed of the current, and the arrow points in the direction of flow. *Source: Karla Panchuk (2002) CC BY 4.0. Based on the exercise described in Chapter 10 of Slingerland & Kump (2011).*

If the model is to be used to track the movement of a pollutant through the lake, it is important to know that it has done a good job of calculating the current velocity. In this case, data from current meters in the lake can be compared to the current velocities that the model calculates (Figure 16.30). The model captures the fact that flow is northward near the margins of the lake, and southward in the middle, but the model current velocities are not exactly the same. This means that the model would do a good job of predicting where the pollution went, but not as good a job at predicting how fast it got there.

Figure 16.30 Comparison of model results with current-meter data. Points plotted beneath the blue dashed line indicate northward flow. Points above indicate southward flow. *Source: Karla Panchuk (2002) CC BY 4.0. Data from Simons and Schertzer (1989).*

The model in this case had a relatively course representation of the lake geography, so a first step would be to make the grid cells smaller to do a better job of simulating the shape of the lake, then look at the flow in finer detail. Another step would be to represent the lake water using a vertical stack of several grid cells to better capture the extent to which bottom friction and wind force affect the lake water at different depths, and to do a better job of representing the water depth.

An Example of Using a Computer Model to Study Past Earth-System Change

The Paleocene-Eocene Thermal Maximum (PETM) was a sudden global warming event that happened approximately 56 million years ago. There was interest in studying this event because its suddenness was thought to be a good analogy for the rapid changes happening in the Earth system today. Cores from ocean-floor rocks show that the oceans became so acidified during the PETM that calcium carbonate sediments dissolved over vast areas of the ocean floor, vanishing entirely from some regions. The same cores also showed a shift in the carbon-isotope composition of calcium carbonate sediments.

Both the acidification and the carbon-isotope shift indicated that a large amount of carbon was added to the Earth system to trigger the PETM. The problem was that there were a number of possible sources for the carbon, and thus a number of possible triggers for the event. Although scientists provided reasoned arguments for their favourite hypotheses, there was no way to know for sure which was the best answer.

To solve this problem, an Earth-system model was used that could test which scenario could best account for the pattern of dissolving calcium carbonate. It took into account the shape of ocean basins, ocean

current circulation, and carbonate system chemistry in ocean water and in sediments. It also took into account changes in sediments once they were deposited.

The steps to using this model were the following:

1. A search was done to locate as many studies of ocean floor sampling sites as possible that had information about changes in the amount of calcium carbonate during the PETM.
2. The model was set up so that it did a good job of reproducing the distribution of calcium carbonate *before* the PETM happened. This was important to ensure that model scenarios began with a realistic set of conditions.
3. Each of the possible scenarios involved carbon coming from different sources in the Earth system, meaning that each scenario could be represented in the model by adding to the atmosphere different amounts of carbon with different carbon isotope compositions. The more carbon a scenario required, the more the calcium carbonate sediments would have dissolved in real life.
4. The model was run for each different amount of carbon. For each scenario, the pattern of calcium carbonate sediments that the model gave was compared to the actual distribution of calcium carbonate sediments known from the data collected in Step 1.

In the end, the model showed that some of the scenarios did not even come close to matching the observations, either dissolving way too much calcium carbonate, or far too little. The model showed that two scenarios did come close to reproducing the pattern of calcium carbonate, and that one did a better job of matching the observations than the other. When it came time to write a report about the experiments, the scientists learned that newly published measurements from another study supported the scenario that the model suggested was best. It would have been acceptable to write a paper describing the model results, and which scenario worked best. However, also being able to comment about new supporting data meant there was a better chance of convincing other scientists that the model results were meaningful.

Predicting the Future of the Earth System with Models

Using models to investigate the Earth system requires careful consideration of how to build the model and run experiments. But it also requires skillful use of real-life measurements to set up the model, and to interpret and evaluate its results. The PETM model study was an example of how a model can be used to test hypotheses about past behaviour of the Earth system. There were data from before, during, and after the event to help set up the model and gauge its effectiveness.

Using Earth-system models to predict the future is a different kind of modeling challenge, because we don't already know what the right answer is. The situation being modeled hasn't happened yet. Scientists who try to predict the future of the Earth system have to do things a bit differently in order to have some confidence in the reliability of their model outcomes:

- They must come up with reasonable forcing scenarios for the model. A model used to predict the future of Earth's climate will need input about what atmospheric greenhouse gas levels will be. That will depend on what actions humans take. Scientists deal with this unknown variable by testing multiple scenarios for greenhouse gas levels, such as what would happen if fossils fuels continue to be used as they have been, or alternatively, what would happen if we completely stopped using fossil fuels tomorrow. The scenarios they choose span a range of possibilities, including extreme cases, to make sure they understand what the possible range of outcomes could be.
- Scientists use some of the data they have to set the model up so that it is a realistic representation of the Earth system at a particular time in history. They then test the model to see if it can reproduce a different set of data later in history. This is a way to see if the model can get the right answer for a time when we know the right answer. Finally, after determining that the model gives reasonable results for times when we know the right answer, it is run for future scenarios.
- To be confident about predictions of future Earth-system change, scientists may collaborate to run their scenarios on many different Earth-system models designed by many different research groups at many different institutions. These models are set up with slightly different mathematical representations of processes, or different levels of detail in geographic representations. The

scientists who built the various models might disagree on what numbers to assign some of the variables, or even which parts of the Earth system are necessary to include. If all the models produce similar results for a particular scenario in spite of representing a wide range of ideas about how such models should work, scientists can be more confident in those results.

- Scientists report **uncertainty** with their model results. It is a common misconception that uncertainty means the same thing as in everyday language—that we just don't know something, or can't say for sure. But for models, uncertainty is a number that indicates the likelihood that a model result is within a certain range of values. It is determined using methods that are themselves the product of careful research. A meaningful discussion of uncertainty will concern a specific model or set of models, a specific variable, and include a specific range of values. It will also include information about how large the uncertainty is compared to the changes they are investigating. If these details are missing from the conversation, it's a clue that "uncertainty" is being used in a common-language way rather than the way that modellers use it. Note that reporting uncertainty is not exclusive to models predicting the future, but it is particularly important for those models because of the great scrutiny Earth-system models receive when they are used to investigate future climate change.

16.5 Humans in the Earth System

The Start of Human Influence on the Earth System

Anthropogenic change in the Earth system is change caused by humans. Many discussions of anthropogenic climate-change place the start of human impacts on the Earth system at the beginning of the industrial era, in the mid 18th century. The industrial era was when humans began to use fossil fuels—at the time, mostly coal—on a much larger scale than before to do things like run manufacturing machinery and trains.

Some climate scientists place the first anthropogenic impacts much earlier, however. Some suggest that anthropogenic climate change began around 8,000 BCE when humans cleared land for agriculture in Europe and the Middle East. Clearing forests for crops is a type of climate forcing because the CO_2 storage capacity of the crops is generally lower than that of the trees they replace. Some climate scientists also point to the creation of wetlands to grow rice in Asia around 5,000 BCE. Creating wetlands is a type of climate forcing because the anaerobic bacterial decay of organic matter within wetlands produces CH_4.

Whether anthropogenic climate change began with the Agricultural Revolution or the Industrial Revolution may be a matter for debate for some, but it is clear that Earth-system change accelerated once the Industrial Revolution began. Part of this is due to the fact that agricultural activities had to be scaled up to feed an ever-growing population. When humans first started growing crops, the world population was approximately 5 million (Figure 16.31), fewer people than live in Toronto today. The world population rose to approximately 18 million when wetland rice cultivation began (fewer people than live within the city limits of Beijing today), to over 800 million at the start of the Industrial Revolution. The world population was estimated at 7,600 million in 2018.

Figure 16.31 World population growth over the past 12,000 years. *Source: Steven Earle (2015) CC BY 4.0. Data from Roser and Ortiz-Ospina (2018)*

The other reason humans accelerated Earth-system change after the start of the industrial era is that human activities required a source of energy, and **fossil fuels** such as coal and oil were that source. Fossil fuels

are those derived largely from plant material that grew, died, and was partially preserved at various times throughout Earth history. The plants removed CO_2 from the atmosphere when they were alive, and stored it in organic compounds in their tissues. The materials accumulated over hundreds of millions of years in settings like swampy forests, shallow seas, and deltas. When fossil fuels are burned, the stored carbon is released back into the atmosphere as CO_2.

The Carbon-Isotope Fingerprints of Fossil Fuel

Carbon isotopes provide insights into the extent to which fossil fuels have impacted the Earth system, because fossil fuels have a unique carbon-isotope fingerprint that is detectable in the atmosphere and in geological materials.

Stable Carbon Isotopes (12-Carbon and 13-Carbon)

When plants transform CO_2 into tissues, the process imparts a unique carbon-isotope signature to the resulting organic matter. Plants preferentially take in CO_2 with the isotope ^{12}C over CO_2 with isotope ^{13}C. They do so in a consistent way, giving plant tissues a distinctive ratio of ^{13}C to ^{12}C. Fossil fuels are derived from plant materials, and they preserve this isotopic ratio.

The ratio of ^{13}C to ^{12}C is commonly expressed relative to a standard to give numbers that are easy to work with and compare. The notation $\delta^{13}C$ refers to the ratio of ^{13}C to ^{12}C in a sample compared to the ratio in a standard, and is expressed in parts per thousand (or per mil, ‰). The standard has a $\delta^{13}C$ of 0‰. Carbon in plant tissues has a $\delta^{13}C$ of -25‰ to -30‰, meaning it has a ^{13}C to ^{12}C ratio that is 25 to 30 parts per thousand lower than the standard. Burning fossil fuel releases CO_2 with that ratio into the atmosphere.

For most of the past 1000 years, the atmosphere has had a $\delta^{13}C$ of approximately -6.5‰. The carbon-isotope composition of organic matter is much lower than that of the atmosphere, so the mixing in of carbon from fossil fuels causes the over-all carbon-isotope composition of the atmosphere to decrease. An analogy for mixing low $\delta^{13}C$ CO_2 into the atmosphere is rapidly adding cold water to a hot bathtub. The faster the cold water is added, the faster the bathwater will cool. The colder the water being added, the faster the bathwater will cool. In this analogy, the atmosphere is the bathtub, and fossil fuels are the water being added. The low $\delta^{13}C$ value of fossil fuels (-25‰ to -30‰) is like very cold water being added.

As we would expect, the carbon isotope composition of the atmosphere takes a sudden downward turn at the same time that humans undertake the Industrial Revolution, and begin burning large quantities of fossil fuels, adding CO_2 to the atmosphere at an accelerating rate (Figure 16.32).

Figure 16.32 A 1000-year record of atmospheric CO_2 levels (blue circles) and carbon-isotope composition (grey circles) measured in Antarctic ice cores. The Industrial Revolution (grey shading), marking the start of the industrial era and the large-scale use of fossil fuels by humans, coincides with a sudden rise in CO_2 levels, and a fall in the carbon-isotope composition of atmospheric CO_2. *Source: Karla Panchuk (2018) CC BY 4.0. Data from Rubino et al (2013).*

Scientists who study past climates on Earth are familiar with carbon-isotope records like this one, because such records are used to reconstruct major changes in the Earth system through their impact on the carbon cycle. In carbon-isotope records from the distant past, a shift of more than 1.5‰ would be enough to catch the attention of a researcher and make them wonder what could have happened.

What is unusual about the 1.5‰ drop today in comparison to those observed in the geological record is how rapidly it is happening. It is more common to see such changes happen over millions of years, not hundreds of years. The rate at which atmospheric CO_2 $\delta^{13}C$ is dropping is approximately 10 times faster than the carbon-isotope shift at the PETM, which is the fastest event ever documented in the rock record.

Carbon dioxide in the atmosphere mixes into the oceans, where organisms take up carbonate ions to make calcium carbonate shells. The 1.5‰ drop has been imprinted in the calcium carbonate of marine organisms like sponges (Böhm et al, 2002), and will remain in the rock record globally, as evidence of human activity. Because of this, and because of many other such markers that are being left in the rock record by human activities (the presence of plastic, for example), some have suggested that it is time to define a new division of geological time, the **Anthropocene Epoch**. The start of the Anthropocene Epoch would mark the point at which human activities became evident in the geological record.

Radioactive Carbon (14-Carbon)

Carbon-14 dating relies on the fact that ^{14}C decays to ^{14}N at a known rate. By knowing the rate, and how much ^{14}C and ^{14}N are present, we can work out how long the decay has been happening. Knowledge of the decay rate of ^{14}C also makes it useful to track fossil fuel additions to the atmosphere.

The rate of decay of a radioactive isotope is expressed as a **half-life**, which in this case is the amount of time it would take half of the ^{14}C atoms in a sample to decay to ^{14}N. The half-life of ^{14}C is 5,730 years. After 10 half-lives, or 57,300 years, there isn't enough ^{14}C left to do an age measurement. Fossil fuels are millions to hundreds of millions of years old, long enough for there to be none of the ^{14}C originally contained by the plant material.

There is a notation system for ^{14}C similar to the $\delta^{13}C$ notation system for the ratio of ^{13}C to ^{12}C, in which the amount of ^{14}C is compared to a standard. Carbon-14 amounts are reported as $\Delta^{14}C$ values in units of ‰. In that system, the atmosphere as a whole had a $\Delta^{14}C$ of 45‰ in 2010, and fossil fuels have a $\Delta^{14}C$ of -1000‰. Effectively, the atmosphere appears to be aging rapidly. In the bathtub analogy for carbon isotopes, adding CO_2 from fossil fuels is like dumping ice into the tub.

The effects of fossil fuel CO_2 on atmospheric $\Delta^{14}C$ levels must account for ^{14}C being made through natural processes in the atmosphere, and decaying away; for the decay of a large pulse of ^{14}C created by nuclear bomb tests; and for other sources of carbon with very low $\Delta^{14}C$ values. Fortunately for scientists tracking fossil fuels by their impact on atmospheric $\Delta^{14}C$, the contribution of low $\Delta^{14}C$ CO_2 from other sources is tiny compared to known rates of fossil fuel emissions, and the other quantities are also well known. Thus, they have been able to determine a decrease in $\Delta^{14}C$ of 3‰ for every 1 ppm of CO_2 added from fossil fuels.

The Carbon Cycle and Change in Today's Earth System

Change in the Earth system is strongly driven by Earth's carbon cycle, the interrelated materials and processes that change carbon from one form to another, and move it from one reservoir to another (Figure 16.33).

Figure 16.33 Flows of carbon in the Earth system. Numbers are rates in billions of tons of carbon (gigatons, Gt) per year. Yellow numbers are rates unrelated to human activity. Red numbers show the contribution of human activities as of 2012. *Source: U.S. Department of Energy (2012) Public Domain*

The CO_2 in the atmosphere is just one part of the carbon cycle. Carbon in the atmosphere is taken in by marine and terrestrial plants, and released when they are decomposed. Microbial activities in the soil and respiration by plants release carbon. Carbon also moves into and out of the ocean through exchange processes at the ocean's surface.

In the carbon cycle today, natural processes as a whole comprise far more of the flow in the carbon cycle than human activities do. For comparison, the relative sizes of flows in Figure 16.33 are illustrated by the size of the arrows. As of 2012, human activities were responsible for approximately 9 billion tons of carbon

added to the atmosphere per year. A large part of the 9 billion tons comes from burning oil, coal, and gas, and some from changes in how land is used (e.g., clearing forests to plant crops, Figure 16.34). Some comes from changes that humans have made that affect the ability of the Earth system to take up carbon.

Figure 16.34 Flow diagram illustrating the pathways through which human activities produce greenhouse gases. The diagram connects the items in each column with flows that ultimately lead to the type of fuel used, and the greenhouse gasses produced. The width of each band is proportional to the quantity flowing from one column to the next. Note that F-Gas refers to anthropogenic fluorinated gases, which are extremely powerful greenhouse gases. *Source: Fischedick et al. (2014), Figure 10.1, based on Bajželj et al. (2013). See Appendix C for terms of use.*

The Earth system has accommodated the 9 billion tons by taking up an additional 3 billion tons per year in photosynthesis, and dissolving an additional 2 billion tons per year in the ocean. The remaining 4 billion tons accumulates in the atmosphere each year because the Earth system does not presently have the capacity to remove it.

The fossil fuels added by humans are particularly problematic because burning them means releasing hundreds of millions of years worth of plant-stored carbon that would otherwise not have been an active part of the carbon cycle today. Contrast this with cutting down a tree and burning the wood. Burning the wood also releases CO_2 from carbon that was stored in plant tissues, but the difference is in timescale and quantity. If a tree grows for 50 years before it is used as fuel, then over a century there is effectively no change in atmospheric CO_2. What carbon the tree took out of the atmosphere decades before, burning and decomposition have returned. (Note: in reality, burning a tree does not have a zero net impact on atmospheric CO_2 levels, because of interactions with other components of the Earth system. In addition, if humans use fossil fuels to power the tools to cut down trees, that will add CO_2 to the atmosphere.)

For fossil fuels, on the other hand, the carbon was removed from the atmosphere tens or even hundreds of millions of years ago. Trees draw down CO_2 before we burn them, balancing out the equation, but with fossil fuels *there is no initial draw-down from our present atmosphere*. Releasing the carbon stored in those fuels results in a net addition to the atmosphere. What makes this even worse is that because fossil fuels have been accumulating for so long, there is an enormous quantity that can be burned. Trees can only be burned as fast as they replace themselves, but with fossil fuels it is like accumulating trees for millions of years, then burning them all at once.

Signals of Present-Day Earth-System Change

Rising Temperatures

From studies of Earth's past climate history, it is clear what to expect as atmospheric CO_2 levels rise. Climate warming is one outcome. We know from ice core records that global average temperatures are warmer now than they have been for most of the last 800,000 years (Figure 16.22). Over the shorter term, direct measurements show that the climate has been on a warming trend after the start of the Industrial Revolution (Figure 16.9). Proxy data making up a revised version of the "hockey stick" diagram—so named because the shape reminded some people of a hockey stick laying on its side—take the record back to 1000 years ago, and show global average temperatures falling until the onset of the industrial era (Figure 16.35).

Figure 16.35 Global average temperature change for the last 1000 years. Blue- The original "hockey stick" diagram showing a reconstruction of northern hemisphere temperatures using tree rings as a proxy. Red- Direct temperature measurements. Green dots- Global temperature reconstruction using a wide range of direct measurements, historical records, and proxies (sediments, ice cores, tree rings, corals, stalagmites, pollen). The original hockey stick diagram was the focus of much controversy because it was the first evidence of anthropogenic climate change that could be understood by the general public. The PAGES2K project sought to bring vast quantities of data to establish once and for all whether a global signal of warming could be reliably discerned. The result was very similar to the original hockey stick. *Source: Karla Panchuk (2018) CC BY-SA 4.0. Modified after Klaus Bittermann (2013) CC BY-SA 4.0*

Sea Level Change

As of April 2018, global sea level has risen approximately 28 cm since 1800. According to satellite data, the average rate of change since 1993 has been a rise of approximately 3 mm per year. Part of the rise is due to the expansion of seawater as it warms. Another part of the rise is from water added by melting glaciers and other year-round land-based snow and ice. Note that melting of sea ice—ice already floating in the ocean—does not contribute directly to sea-level rise because the ice is already floating in the ocean.

Based on how much melting has occurred thus far, sea levels are projected to rise to between 47 cm and 130 cm above 1880 levels (Figure 16.36). However, there is some uncertainty about how melting rates will respond to changes in the Earth system that result from climate change, such as changes in currents, or seawater beneath the leading edge of melting ice sheets warming the ice from beneath. With that uncertainty factored in, sea level rise could be as low as 33 cm above 1880 levels, or more than 2 m higher.

Figure 16.36 Measured and projected change in global average sea level. Data come from proxy records as well as from direct measurements from tidal gauges and satellite data. Projected sea level rise could be as little as 33 cm over 1800 levels, or as much as 206 cm. *Source: Karla Panchuk (2018) modified after Steven Earle (2015) CC BY 4.0 and J. Willis, Jet Propulsion Laboratory (2013). See Appendix C for terms of use.*

Keep in mind that the global average is indeed an average. Where ocean waters experience more warming, and thus more thermal expansion, sea level rise may be greater than elsewhere. Regions that are rebounding as ice melts could experience less sea level rise, or even a fall in sea level, because the elevation of the terrain is actually increasing over time. On the other hand, regions on the peripheral bulge around the margins of ice sheets could experience greater than average sea level rise because the terrain will subside at the same time that the oceans are gaining volume.

Areas that become flooded could experience greater than average sea level rise, because the weight of water causes the land to subside further. In the aftermath of Hurricane Harvey in September of 2017, measurements were reported that showed subsidence of up to 1.5 cm in the region of Houston, Texas. In this case, some of the subsidence could have been from sediments being compressed under the weight of flood waters, however the weight of water, like the weight of ice, does cause the crust to float lower in the mantle.

Melting Ice Sheets

Keeping track of how rapidly ice sheets are melting is important both for being able to predict future sea level change, and for knowing in general how rapidly the Earth system is changing. In a recent study, Bamber et al. (2018) analyzed satellite measurements to determine how much mass had been lost from the Antarctic ice sheets, the Greenland Ice Sheet, and from other glaciers and ice caps around the world since 1992 (Figure 16.37).

Figure 16.37 Large ice sheets of Antarctica and Greenland (blue) and glaciers and ice caps (yellow). Circles are proportional to the area of each region that is covered by glaciers. The green part of the circle indicates the proportion of the ice with margins resting on land, and the blue part indicates margins in the ocean. This difference is important in part because of the potential for faster melting when the base of an ice sheet is in contact with warming seawater. *Source: Bamber et al. (2018) CC BY 4.0.*

The study found that over all, the mass of ice in ice sheets, ice caps, and glaciers has been falling at an increasing rate since 1992, and therefore adding to sea level at an increasing rate (Figure 16.38). The exception is the East Antarctic Ice Sheet, which actually showed an increase in mass during the studied interval. This is because snowfall has increased in the East Antarctic, to the point where more snow is falling now than at any time in the past 2000 years (Medley et al., 2017). The East Antarctic is warming just as the West is, but the difference is that the winds that preferentially bring precipitation to the East rather than the West can carry more moisture because the air is warmer.

Figure 16.38 Results of a study of the change in mass of ice on Earth's surface. Satellite data show that over all, melting has accelerated since 1992. *Source: Source: Karla Panchuk (2018) CC BY 4.0, modified after Bamber et al. (2018) CC BY 4.0.*

16.6 Welcome to the Anthropocene

Geologists still have not agreed on an official start date for the Anthropocene Epoch, but it is unlikely that they would disagree that we are presently in it. What this means in practical terms is that humans today are experiencing the results of past human influence on the Earth system, and humans in the future will experience the results of decisions made today.

Data on Key Sources of Radiative Forcing

The **Intergovernmental Panel on Climate Change** (IPCC) was established by the United Nations in 1988 to help with those decisions. Is responsible for reviewing the scientific literature on climate change and issuing periodic reports on several topics, including the scientific basis for understanding climate change, our vulnerability to observed and predicted climate changes, and what we can do to limit climate change and minimize its impacts.

In their Fifth Assessment Report, Summary for Policy Makers, the IPCC identified the main contributions to heating or cooling our atmosphere, and the impact they had in 2011 compared to 1750, before the industrial age (Figure 16.39). The main *natural* source of radiative forcing is the sun, but it contributes relatively little compared to anthropogenic sources. The greatest contribution over all comes from anthropogenic CO_2 (top row). Note that the bar sizes indicate the relative contribution, and correspond to the scale at the bottom of the diagram. Bars to the left of the zero line indicate cooling of the atmosphere, and bars to the right indicate warming. The column on the far right indicates how confident scientists are in their assessment of radiative forcing for each source (VH = very high; H = high; M = medium; L = low).

Figure 16.39 Anthropogenic and natural contributions to radiative forcing in 2011 compared to 1750. *Source: IPCC (2013). See Appendix C for terms of use.*

The biggest anthropogenic contributor to warming, CO_2, accounts for 50% of positive forcing. CH_4 and its atmospheric derivatives (CO_2, H_2O, and O_3) account for 29%, and the halocarbon gases (mostly leaked from air-conditioning appliances) and nitrous oxide (N_2O, from burning fossils fuels) account for 5% each. Carbon monoxide (CO, also produced by burning fossil fuels) accounts for 7%, and the volatile organic compounds other than methane (NMVOC) account for 3%.

CO_2 emissions come mostly from coal- and gas-fired power stations, motorized vehicles (cars, trucks, and aircraft), and industrial operations (e.g., smelting), and indirectly from forestry. CH_4 emissions come from production of fossil fuels (escape from coal mining and from gas and oil production), livestock farming (mostly beef), landfills, and wetland rice farming. N_2O and CO come mostly from the combustion of fossil fuels. In summary, close to 70% of current anthropogenic GHG emissions come from fossil fuel production and use, while most of the rest comes from agriculture and landfills.

Projected Warming

The IPCC projects future warming by using scenarios that represent different possible paths that take into account global population, economic development, and global co-operation. In their best-case scenario, called RCP2.6 (RCP stands for representative concentration pathway, and 2.6 refers to a radiative forcing of 2.6 W/m^2), global CO_2 emissions begin to decline after 2020, and go to zero by 2080. Their worst-case scenario is RCP8.5, in which global CO_2 emissions continue to rise rapidly. Intermediate scenarios reflect cases where global CO_2 emissions eventually peak, but do not go to zero.

Figure 16.40 Model projections to 2100 for surface temperatures (a) and the extent of sea ice in September in the northern nemisphere (b). Black line: data; Grey shading: model attempts to simulate conditions from 1950 to 2005; Blue: best-case scenario with peak CO_2 emissions in 2020 and zero emissions by 2080; Red: worst-case scenario with no decline in emissions. Numbers indicate the number of models run for each scenario. Shading indicates range of uncertainty. *Source: IPCC (2013). See Appendix C for terms of use.*

In the best-case scenario, temperatures stabilize at approximately 1 ºC above 1996 temperatures by the end of the century. As of 2018 there is no global plan in place that will meet the timeline of RCP2.6, and cause emissions to decline after 2020. For sea ice, all but the best-case scenario have the average sea ice extent beneath the dashed line, which is effectively means the northern hemisphere would be ice-free.

Extreme Climate-System Events

In 2015 the World Meteorological Organization (WMO) published an analysis of extreme climate-system events, and showed that the number of such events globally was almost 5 times higher in the decade 2001-2010 than in the decade 1971-1980 (Figure 16.41).

Figure 16.41 Numbers of climate-system disasters between 1971 and 2010. *Source: WMO (2015) Atlas of Mortality and Economic Losses from Weather, Climate and Water Extremes: 1970-2012. See Appendix C for terms of use.*

Analyses of natural disasters and extreme weather events by the European Academies' Science Advisory Council (EASAC) showed that from 1980 to 2010, extreme weather, floods, and mass movement related to floods increased by more than three-fold in Europe. In the same study, analyses of insurance-industry data showed that globally, Europe and South America are the *least* affected. The number of events has been more than twice as great in Asia, Australia/Oceania, and North America (EASAC, 2013). The EASAC recently published an update to their study, to include data up to 2016. They now say that floods and mass movement events have increased in Europe by more than 4 times from 1980 levels (EASAC, 2018).

Extreme Temperatures

The major types of climate-related disasters reported by the WMO report are floods and storms, but the health implications of extreme temperatures are also becoming a great concern. In the decade 1971 to 1980, extreme temperatures were the fifth most common natural disasters; by 2001 to 2010, they were the third most common.

For several weeks in July and August of 2010, a massive heat wave affected western Russia, especially the area southeast of Moscow. Temperatures soared to over 40°C, as much as 12°C above normal over a wide area, and wildfires raged in many parts of the country. Over 55,000 deaths are attributed to the heat and to respiratory problems associated with the fires (Figure 16.42).

Figure 16.42 Temperature anomalies across Russia and neighbouring regions during July 2010. *Source: Earth Observatory (2010) Public Domain*

A clue to explaining the heatwave was simultaneous flooding in Pakistan: in July, rainfalls were 70% higher than normal, and in August rainfalls were 102% above normal. Flooding from heavy rains combined with a burst dam impacted 18 million people, killing 1,985. These events were linked by a change in atmospheric circulation that caused weather patterns to stall out over Russia and Pakistan. The change itself was due in part to a decrease in the temperature differences between high and low latitudes, caused by warming in the north.

Sea-Level Rise and Tropical Storm Impacts

Projections for sea-level rise to the end of this century vary widely (Figure 16.36), however it is already the case that low-lying coastal regions are experiencing increased flooding during storm events. Flood defenses designed decades ago are or will soon be insufficient in the face of increasingly frequent and more extreme storm events. In a 2008 report, the Organisation for Economic Co-operation and Development (OECD) estimated that by 2070 approximately 150 million people living in coastal areas could be at risk of flooding due to the combined effects of sea-level rise, increased storm intensity, and land subsidence. The assets at risk (buildings, roads, bridges, ports, etc.) are on the order of $35 trillion ($35,000,000,000,000). Countries with the greatest exposure of population to flooding are China, India, Bangladesh, Vietnam, U.S.A., Japan, and Thailand. Some of the major cities at risk include Shanghai, Guangzhou, Mumbai, Kolkata, Dhaka, Ho Chi Minh City, Tokyo, Miami, and New York.

One of the other risks for coastal populations, besides sea-level rise, is that climate warming is also associated with an increase in the intensity of tropical storms (e.g., hurricanes or typhoons), which almost always bring serious flooding from intense rain and storm surges. It is difficult for climatologists to ascribe all of the characteristics of a particular tropical storm to climate change, but it is much easier to make a case for why tropical storms in general would become stronger, and have more of an impact on coastal regions.

When Hurricane Sandy hit New Jersey and New York in 2012 (Figure 16.43), it had a devastating impact in part because it smashed into the coast directly rather than moving parallel to the coast. There was some speculation that the unusual path was the result of changes in atmospheric circulation that were related to climate change. We can't say for certain whether climate change caused Sandy to move the way it did, but we can say with confidence that climate change worsened the impact of Sandy through sea-level rise. Consider that, over the previous 100 years, sea level had risen a foot at the southern-most tip of Manhattan Island. The island was hit by a 14-foot storm-surge, higher than it would have been without sea-level rise.

Figure 16.43 Damage from Hurricane Sandy on the coast of New Jersey. *Source: Master Sgt. Mark C. Olsen, New Jersey National Guard (2010) Public Domain*

Tropical Storm Intensity

Tropical storms get their energy from the evaporation of warm seawater in tropical regions. In the Atlantic Ocean, this takes place between 8° and 20° N in the summer. A correlation has been observed between variations in the sea-surface temperature (SST) of the tropical Atlantic Ocean (Figure 16.44, red) and the amount of power represented by Atlantic hurricanes between 1950 and 2008 (blue). Not only has the overall intensity of Atlantic hurricanes increased with the warming since 1975, but the correlation between hurricanes and sea-surface temperatures is very strong over that time period.

Figure 16.44 Relationship between Atlantic tropical storm cumulative annual intensity and Atlantic sea-surface temperatures. *Source: Steven Earle (2015) CC BY 4.0*

Because warm air is able to hold more water than cold air, the general global trend over the past century has been one of increasing precipitation (Figure 16.45). Note that this does not mean that precipitation has increased everywhere in the world. It means that even taking into account regions that have become drier, there is still more precipitation over all.

Figure 16.45 Global precipitation from 1901 to 2000. *Source: Karla Panchuk (2018) CC BY 4.0, with data from NASA/GISS*

Thawing Permafrost

Most of northern Canada has a layer of permafrost that ranges from a few centimetres to hundreds of metres in thickness. The same is true in Alaska, Russia, and Scandinavia. One problem that northern communities are currently dealing with is that thawing permafrost is weakening sediments that were previously solid and stable. This can cause structures to settle, or even to sustain serious damage. Another problem with the weakening that occurs when permafrost thaws is that mass wasting can occur (Figure 16.46). This is particularly dangerous for northern communities along coastlines, where homes may be near the shore and the slopes beneath them further weakened by strong wave action during severe storms.

Figure 16.46 A degrading permafrost site on the north coast of Alaska. *Source: Alaska Science Center, U. S. Geological Survey (2016) Public Domain*

A longer-term issue with thawing permafrost is that micro-organisms begin to decompose organic matter within the newly thawed sediments, releasing CO_2 and CH_4. The amount of these greenhouse gases released by decomposing organic matter could be enough to generate a significant positive feedback, accelerating warming further. In some polar regions, including northern Canada, permafrost also includes methane hydrate, a highly concentrated form of CH_4 trapped in solid form. Breakdown of permafrost releases this CH_4 as well.

Pests and Disease

The geographical ranges of diseases and pests, especially those caused or transmitted by insects, have been shown to extend toward temperate regions because of climate change. West Nile virus and Lyme disease are two examples that already directly affect Canadians, while dengue fever could be an issue in the future. Canadians are also indirectly affected by the increase in populations of pests such as the mountain pine beetle (Figure 16.47).

Figure 16.47
Mountain pine beetle damage in Manning Park, British Columbia.
Source: Jonhall (2010)
CC BY 3.0

Chapter 16 Summary

The topics covered in this chapter can be summarized as follows:

16.1 What Is the Earth System?

Viewing Earth as a system allows us to take into account the complex ways in which the atmosphere, hydrosphere, biosphere, and lithosphere interact. Positive feedbacks amplify changes in the Earth system, and negative feedbacks reduce them. The stability of the Earth system will depend on what feedbacks are available. The presence of ice sheets makes the Earth system less stable.

16.2 Causes of Climate Change

Weather describes day-to-day conditions, but climate refers to the long-term average conditions over decades or longer. Climate forcings alter climate. They include processes that change the rate and location of solar energy reaching Earth's surface; processes that alter how ocean currents move heat around Earth's surface; and processes that affect how heat moves into and out of the atmosphere. Climate forcings operate on a range of timescales, from billions of years to less than a decade. Changes in greenhouse gas concentrations and albedo are two climate forcings affected by human activities.

16.3 Methods for Studying Past Climate

Climate conditions for some of human history can be determined from direct measurements that have been recorded, but for studying paleoclimate it is often necessary to use proxy data. Proxy data come from natural materials that behave in a systematic way in response to climate conditions like temperature or

precipitation. Proxies include tree ring data, stable isotopes, measurements of gas bubbles trapped in ice, and the geographic distribution of rocks and fossils.

16.4 Computer Models of the Earth System

Earth-system models use mathematical equations to simulate Earth-system processes. Models are set up and checked using real-life measurements. Model uncertainty is a number that tells us the likelihood that a particular model result falls within a certain range of values. It is a way to evaluate whether results can be used to draw meaningful conclusions.

16.5 Humans in the Earth System

Data show recognizable anthropogenic influence on the Earth system beginning when humans began to use fossil fuels for industrial purposes. CO_2 in the atmosphere has the isotopic fingerprints of fossil fuels. The flow of anthropogenic carbon into the Earth system is relatively small compared to some natural flows, but natural processes do not remove all of what humans put in, causing CO_2 to accumulate in the atmosphere.

16.6 Welcome to the Anthropocene

Humans today are experiencing the results of past human influence on the Earth system, and humans in the future will experience the results of decisions made today. The main source of radiative forcing is anthropogenic CO_2. Humans are already experiencing extreme climate events related to warming. The best-case projected warming scenario will stabilize global average temperatures at ~1ºC above 1996 temperatures by the end of the century, but this will require a peak in anthropogenic CO_2 emissions by 2020, and an end to anthropogenic CO_2 emissions by 2080.

Review Questions

1. If you receive positive feedback on a project, this means someone says you did a good job. Does this mean a positive feedback in the climate system is also something good? Explain.
2. Why does the presence of ice sheets cause the Earth system to be less stable?
3. Using the orbital information on eccentricity, tilt, and precession, we could calculate variations in insolation for any latitude on Earth and for any month of the year. Why do we focus on a latitude of 65° N during September?
4. Explain how the positioning of Gondwana at the south pole contributed to glaciation during the Paleozoic.
5. If the major currents in the oceans were to slow down or stop, how would that affect the distribution of heat on Earth, and what effect might that have on glaciation?
6. Most volcanic eruptions lead to short-term cooling, but long-term sustained volcanism can lead to warming. Describe the mechanisms for these two different consequences.
7. What property of greenhouse gases allows them to absorb infrared radiation and thus trap heat within the atmosphere?
8. What are the advantages of proxy data over direct measurements of climate? What are the disadvantages of using proxy data?
9. What are some steps that scientists take to help ensure that meaningful conclusions can be drawn from Earth-system computer models?
10. What evidence tells us that rising atmospheric CO_2 levels are primarily the result of humans burning fossil fuels?
11. The flows of carbon through the carbon cycle that are related to human activities are actually much smaller than some natural flows. Why do human activities have such a large impact on the carbon cycle?
12. Some people argue that climate has always changed, and present-day climate change is not significant compared to the long-term record of Earth's climate. Is it true that present-day climate

signals show that changes are not significant compared to Earth's long-term climate history? Explain your answer.

13. What climate mechanism links both extreme heat waves and droughts, and extreme flooding and rainfall events to climate change?
14. What evidence suggests that increasingly intense tropical storms can be expected as climate warms?
15. What are the potential impacts of thawing permafrost?

Answers at the end of the chapter

Resources

Canadian Climate Normals: http://climate.weather.gc.ca/climate_normals/index_e.html

See the online version of the textbook for other data sources.

References

Bajželj, B., Allwood, J. M., and Cullen, J. M. (2013). Designing Climate Change Mitigation Plans That Add Up. *Environmental Science & Technology 47*, 8062-8069. doi: 10.1021/es400399h. Retrieved from http://pubs.acs.org/doi/pdf/10.1021/es400399h

Bamber, J. L., Westaway, R. M., Marzeion, B., & Wouters, B. (2018). The land ice contribution to sea level during the satellite era. *Environmental Research Letters 13*(2018). https://doi.org/10.1088/1748-9326/aac2f0

Berwyn, B. (2017, 27 March). *Climate Change-Fueled Jet Stream Linked to Brutal Floods and Heatwaves, Says Study*. Retrieve from https://insideclimatenews.org/news/27032017/climate-change-global-warming-extreme-weather-jet-stream-michael-mann-penn-state

Böhm, F., Haase-Schramm, A., Eisenhauer, A., Dullo, W.-C., Joachimski, M. M., Lehnert, H., & Reitner, J. (2002). Evidence for preindustrial variations in the marine surface water carbonate system from coralline sponges. *Geochem. Geophys. Geosyst., 3*(3), 10.1029/2001GC000264. Retrieved from http://onlinelibrary.wiley.com/doi/10.1029/2001GC000264/epdf

Caesar, I., Rahmstorf, S., Robinson, A., Feulner, G., & Saba, V. (2018). Observed fingerprint of a weakening Atlantic Ocean overturning circulation. *Nature 556*. 194-196. https://doi.org/10.1038/s41586-018-0006-5

Charley, S. (2013, 13 June). *Hurricane Sandy took highly unusual path, but climate change doesn't get the blame – yet.* Retrieved from https://blogs.agu.org/geospace/2013/06/03/hurricane-sandy-took-highly-unusual-path-but-climate-change-doesnt-get-the-blame-yet/

Earth System Research Laboratory, Global Monitoring Division, NOAA (n.d.). *The Data: What 14C Tells Us.* Retrieved from https://www.esrl.noaa.gov/gmd/outreach/isotopes/c14tellsus.html

European Academies' Science Advisory Council. (2013). *Trends in Extreme Weather Events in Europe: Implications for National and European Union Adaptation Strategies.* Halle (Saale), Germany: German National Academy of Sciences. Retrieved from https://easac.eu/publications/details/trends-in-extreme-weather-events-in-europe/

European Academies' Science Advisory Council. (2018). *Extreme weather events in Europe. Preparing for climate change adaptation: An update on EASAC's 2013 study.* Halle (Saale), Germany: German National Academy of Sciences. Retrieved from https://www.knaw.nl/shared/resources/internationaal/bestanden/EASACStatement_Extreme_Weather_EventsMarch2018.pdf

Fischedick M., Roy, J., Abdel-Aziz, A., Acquaye, A., Allwood, J. M., Ceron, J.-P., Y. Geng, Y., Kheshgi, H., Lanza, A., Perczyk, D., Price, L., Santalla, E., Sheinbaum, C., and Tanaka, K. (2014). Industry. In: *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickemeier, P., Kriemann, B., Savolainen, J., Schlömer, S., von Stechow, C., Zwickel, T., and Minx, J. C. (eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. Retrieved from https://www.ipcc.ch/pdf/assessment-report/ar5/wg3/ipcc_wg3_ar5_chapter10.pdf

Furphy, D. (2013, 28 September). *What on earth is an RCP? A quick guide to carbon dioxide emissions scenarios used by the IPCC Assessment Report 5.* Retrieved from https://medium.com/@davidfurphy/what-on-earth-is-an-rcp-bbb206ddee26

Ingleby, B., & Huddleston, M. (2007). Quality control of ocean temperature and salinity profiles—Historical and real-time data. *Journal of Marine Systems, 65,* 158-175. doi:10.1016/j.jmarsys.2005.11.019

Intergovernmental Panel on Climate Change. (2013). *Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel*

on Climate Change [Stocker,T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. Retrieved from https://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5_SPM_FINAL.pdf

Jouzel, J., Masson-Delmotte, V., Cattani, O., Dreyfus, G., Falourd, S., Hoffmann, G., Minster, B., Nouet, J., Barnola, J.M., Chappellaz, J.A., Fischer, H., Gallet, J.C., Johnsen, S.J., Leuenberger, M., Loulergue, L., Luethi, D., Oerter, H., Parrenin, F., Raisbeck, G.M., Raynaud, D., Schilt, A., Schwander, J., Selmo, E., Souchez, R., Spahni, R., Stauffer, B., Steffensen, J.P., Stenni, B., Stocker, T.F., Tison, J.L., Werner, M., & Wolff, E.W. (2007). Orbital and Millennial Antarctic Climate Variability over the Past 800,000 Years. *Science 317*(5839), 793-797. Get data: https://www.ncdc.noaa.gov/paleo-search/study/6080

Lüthi, D., Le Floch, M., Bereiter, B.; Blunier, T., Barnola, J.M., Siegenthaler, U., Raynaud, D., Jouzel, J., Fischer, H., Kawamura, K., & Stocker, T.F. (2008). High-resolution carbon dioxide concentration record 650,000-800,000 years before present. Nature (453), 379-382. Get data: https://www.ncdc.noaa.gov/paleo-search/study/6091

Madrigal, A. C. (2017, September 5) *The Houston Flooding Pushed the Earth's Crust Down 2 Centimeters.* Retrieved from https://www.theatlantic.com/technology/archive/2017/09/hurricane-harvey-deformed-the-earths-crust-around-houston/538866/

Medley, B., McConnell, J. R., Neumann, T. A., Reijmer, C. H., Chellman, N., Sigl, M., & Kipfstuhl, S. (2018). Temperature and Snowfall in Western Queen Maud Land Increasing Faster Than Climate Model Projections. *Geophysical Research Letters 45*(3), 1472-1480. https://doi.org/10.1002/2017GL075992

Met Office Hadley Centre (n.d.) *EN3: quality controlled subsurface ocean temperature and salinity data.* Retrieved from https://www.metoffice.gov.uk/hadobs/en3/

Rahmstorf, S. (2013). *Most Comprehensive Paleoclimate Reconstruction Confirms Hockey Stick.* Retrieved from https://thinkprogress.org/most-comprehensive-paleoclimate-reconstruction-confirms-hockey-stick-e7ce8c3a2384/

RAPID-AMOC (n.d.) *RAPID: monitoring the Atlantic Meridional Overturning Circulation at 26.5°N since 2004.* Retrieved from http://www.rapid.ac.uk/

Ridge, J.C. (2008) The North American Glacial Varve Project. Retrieved from http://eos.tufts.edu/varves

Robson, J., Hodson, D., Hawkins, E., & Sutton, R. (2014). Atlantic overturning in decline? *Nature Geoscience 7*(2-3). https://doi.org/10.1038/ngeo2050

Roser, M., & Ortiz-Ospina E. (2018). *World Population Growth.* Retrieved from https://ourworldindata.org/world-population-growth

Rubino, M., Etheridge, D. M., Trudinger, C. M., Allison, C. E., Battle, M. O., Langenfelds, R. L., Steele, L. P., Curran, M., Bender, M., White, J. W. C., Jenk, T. M., Blunier, T., & Francey, R. J. (2013). A revised 1000 year atmospheric $\delta^{13}C$-CO_2 record from Law Dome and South Pole, Antarctica. *Journal of Geophysical Research: Atmospheres 118,* 8482–8499. doi:10.1002/jgrd.50668

Scott, M. (2011, 6 April). *Heavy Rains and Dry Lands Don't Mix: Reflections on the 2010 Pakistan Flood.* Retrieved from https://earthobservatory.nasa.gov/Features/PakistanFloods

Shaffrey, L.C., Stevens, I., Norton, W. A., Roberts, M. J., Vidale, P. L., Harle, J. D., Jrrar, A., Stevens, D. P., Woodage, M. J., Demory, M. E., Donners, J., Clark, D. B., Clayton, A., Cole, J. W., Wilson, S. S., Connelley, W. M., Davies, T. M., Iwi, A. M., Johns, T. C., King, J. D., New, A. L., Singlo, J. M., Slingo, A., Steenman-Clark, L., & Martin, G. M. (2009). U.K. HiGEM: The New U.K. High-Resolution Global Environment Model—Model Description and Basic Evaluation. *Journal of Climate 22*, 1861 - 1896. doi:10.1175/2008JCLI2508.1

Simons, T. J., & Schertzer, W. M. (1989) *The circulation of Lake Ontario during the summer of 1982 and the winter of 1982/83.* Burlington, ON: Environment Canada.

Slingerland, R., & Kump, L. (2011). *Mathematical Modeling of Earth's Dynamical Systems: A Primer.* Princeton, NJ: Princeton University Press.

Smith, D. M., & Murphy, J. M. (2007). An objective ocean temperature and salinity analysis using covariances from a global climate model. *Journal of Geophysical Research, 112*(C02022). doi:10.1029/2005JC003172

Thornalley, D. J. R., Oppo, D. W., Ortega, P., Robson, J. I., Brierley, C. M., Davis, R., Hall, I. R., Moffa-Sanchez, P., Rose, N. L., Spooner, P. T., Yashayaev, I., & Keigwin, L. D. (2018). Anomalously weak Labrador Sea convection and Atlantic overturning during the past 150 years. *Nature 556,* 227-230. https://doi.org/10.1038/s41586-018-0007-4

World Meteorological Organization. (2014). *Atlas of Mortality and Economic Losses from Weather, Climate, and Water Extremes (1970-2012).* Geneva, Switzerland: World Meteorological Association. Retrieved from https://public.wmo.int/en/resources/library/atlas-mortality-and-economic-losses-weather-and-climate-extremes-1970-2012

Zhang, J., & Crowley, T. (1989). Historical Climate Records in China and Reconstruction of Past Climates. *Journal of Climate 2*(8), 833-849. https://doi.org/10.1175/1520-0442(1989)002<0833:HCRICA>2.0.CO;2

Answers to Review Questions

1. A positive feedback does not mean that something good is happening in the Earth system. Rather, it means that a change is being amplified by other Earth-system processes that were triggered by the change. From a human perspective, having a change amplified could be good or bad.

2. When ice is present, relatively small changes in temperature can trigger albedo feedbacks that amplify the change. If no ice is present, a small increase in temperature might not have much of an impact. But if ice is present, that small increase could be enough to cause the ice to melt, and decrease Earth's albedo. If ice is created when the climate cools, further cooling can be triggered if the ice increases Earth's albedo significantly.

3. We use 65° for estimating the glaciation potential of orbital variations because glaciers are most likely to form at high latitudes, where temperatures are cooler, and insolation less direct. We use 65° N rather than 65° S because for more than 50 million years the continents have been concentrated in the northern hemisphere. We use July instead of January, because for ice to form and remain year-round, it is necessary to have cool summers.

4. Gondwana was situated over the south pole for much of the Paleozoic. Not only was the land subject to temperatures cooler over all, and less direct insolation for part of the year, but the large size of the continent limited the extent to which heat from the ocean could reach the interior of the continent. The development of ice on Gondwana further cooled the planet through albedo feedback.

5. If the major currents in the oceans were to slow down or stop, the tropics would get hotter and the high-latitude areas would get colder, leading to expansion of glaciers and sea ice. The various feedbacks (e.g., higher albedo because of increased ice cover) would result in a cooler climate over all.

6. From a climate perspective, the two important volcanic gases are SO_2 and CO_2. SO_2 is converted to aerosols which block sunlight and can lead to short-term cooling (lasting years). CO_2 can lead to warming, but only in situations where there is an elevated level of volcanism over the long term.

7. Greenhouse gases vibrate at frequencies that are similar to those of infrared radiation. When infrared radiation hits a greenhouse gas molecule, the molecule’s vibrational energy is enhanced and the radiation energy is converted into heat, which is trapped within the atmosphere.

8. Proxy data can provide information about climate in the distant past, for times during which there were no direct measurements. As new techniques are developed, analyses of proxy data can improve, whereas direct measurements will have experimental errors related to the instruments used to make them at the time. A disadvantage of proxy data is that we don't always have a detailed timeline to go with the data, and may need to rely on other lines of evidence to understand when the proxy data should apply. Special equipment or techniques may be needed to get proxy data from geological materials. It is also necessary to ensure that geological processes have not altered the materials in some way that makes the results of analyses unreliable.

9. Real-life measurements are used to set up the model, and to test that it can provide realistic results. Many different models may be used to study the same scenario, and the results compared. This helps scientists to compare the results of representing the Earth system in different ways. An estimate of uncertainty is provided so scientists can decide whether or not the range of uncertainty is too big, given the size of the Earth-system change being studied.

10. Data from ice cores and direct measurements show that atmospheric CO_2 levels began to rise very rapidly at the same time that humans began using fossil fuels extensively to power activities like manufacturing and transportation. At the same time, the stable carbon-isotope composition of atmospheric CO_2 began to decrease at a rate that is consistent with the stable carbon-isotope fingerprint of plant-derived organic matter. Also simultaneously, the radiocarbon age of atmospheric CO_2 began to fall, consistent with carbon from very ancient plant sources being added to the atmosphere. Aside from the timing of geochemical changes matching the onset of the industrial era, there are no other sources of carbon that can account for the geochemical fingerprint in the atmosphere.

11. The reason that relatively small anthropogenic flows of carbon into the carbon cycle can have a big impact is that they aren't being balanced by natural processes that can remove the extra carbon. The result is that carbon from anthropogenic activities is accumulating in the atmosphere and ocean.

12. In the past, Earth's climate has been both much colder and much hotter than today. However, present day temperatures are significantly different compared to global average temperatures over the past 1000 years, and stand out in the data from the past 800,000 years. Atmospheric CO_2 levels are much higher than at any time during the past 800,000 years, and on that timescale, the increase is almost instantaneous. Therefore, the magnitude of present-day change in the Earth-system stands out clearly from background conditions predating the appearance of anatomically modern

humans. It is also happening much faster than even the most rapid carbon-release event that we know of in the rock record.

13. Climate warming happens more rapidly at high latitudes more than low latitudes, decreasing the temperature difference between them. The decreased difference allows atmospheric circulation patterns to "stall out" over a location for longer than usual. What would normally be a short heatwave lasts much longer, and a precipitation event that would normally move elsewhere after a short time is now halted in place for weeks.

14. See Figure 16.44. The power of tropical storms depends on the warmth of seawater. Data show that during particularly warm intervals, tropical storms are stronger. Data also show that the strength of storms over all is increasing as sea surface temperatures rise over time.

15. Frozen soils provide a solid foundation for buildings in northern communities. Melting of permafrost means the soil beneath buildings is weakened, and buildings can suffer structural damage. Thawing means that hillsides are more prone to mass wasting, especially terrain along coasts. In terms of the climate system, thawing of permafrost allows micro-organisms to decompose organic matter within previously-frozen sediments, releasing additional CO_2 and CH_4, and contributing to a positive feedback.

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-NC-SA 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 17. Glaciation

Adapted by Joyce M. McBeth from *Physical Geology* by Steven Earle

Learning Objectives

After reading this chapter, completing the exercises within it, and answering the questions at the end, you should be able to:

- Define, draw, and describe the major features of glaciers
- Explain the differences between continental and alpine glaciation
- Summarize how snow and ice accumulate above a glacier's equilibrium line and are converted to ice
- Explain how basal sliding and internal flow facilitates the movement of ice from the upper part to the lower part of an alpine glacier
- Describe and identify the various landforms related to alpine glacial erosion, including U-shaped valleys, arêtes, cols, horns, hanging valleys, truncated spurs, drumlins, roche moutonnées, glacial grooves, and striae
- Identify various types of glacial lakes, including tarns, finger lakes, moraine lakes, and kettle lakes
- Describe the nature and origin of lodgement till and ablation till
- Describe the nature and origin of glaciofluvial, glaciolacustrine, and glaciomarine sediments
- Describe the timing and extent of Earth's past glaciations, going as far back as the early Proterozoic
- Describe the important geological events that led up to the Pleistocene glaciations
- Explain how the Milankovitch orbital variations along with positive climate feedback mechanisms may have controlled the timing of the Pleistocene glaciations

Figure 17.1 Glaciers in the Alberta Rockies: Athabasca Glacier (centre left), Dome Glacier (right), and the Columbia Icefield (visible above both glaciers). The Athabasca Glacier has prominent lateral moraines on both sides. *Source: Steven Earle (2015) CC BY 4.0. https://opentextbc.ca/geology/*

A **glacier** is a long-lasting (decades or more) body of ice that is large enough to move under its own weight. They are at least tens of metres thick and at least hundreds of metres in extent. About 10% of Earth's land surface is currently covered with glacial ice, and although the vast majority of this is in Antarctica and Greenland, there are many glaciers in Canada, especially in the mountainous parts of BC, Alberta, and the Yukon, and in the far north (Figure 17.1). At various times during the past million years, glacial ice has been much more extensive, covering at least 30% of the Earth's land surface at times.

Glaciers currently represent the largest repository of fresh water on Earth (~69% of all fresh water). They are highly sensitive to changes in climate, and in recent decades have been melting rapidly worldwide (Figure 17.2). Although some of the larger glacial masses may still last for several centuries, smaller glaciers, including many in western Canada, may be gone within decades. For mountainous regions, glaciers are an important sources of drinking water. Rapid glacial melting is a troubling issue for western Canadians because glacial ice is an important part of the hydrologic cycle in glaciated regions. Irrigation systems in BC, and across Alberta and Saskatchewan, are replenished by meltwater originating from glaciers in the Coast Range and the Rocky Mountains.

Figure 17.2 Example of rapid melting of a glacier over a 63-year period. Muir Glacier, Alaska. *Source: NASA.Public Domain. https://climate.nasa.gov/climate_resources/4/.*

Figure 17.3 Part of the continental ice sheet in Greenland, with some outflow alpine glaciers in the foreground. *Source: Steven Earle (2015) CC BY 4.0. https://opentextbc.ca/geology/*

17.1 Types of Glaciers

There are two main types of glaciers: continental glaciers and alpine glaciers. Latitude, topography, and global and regional climate patterns are important controls on the distribution and size of these glaciers.

17.1.1 Continental Glaciers

Continental glaciers cover vast areas of land. Today, continental glaciers are only present in extreme polar regions: Antarctica and Greenland (Figure 17.3). Historically, continental glaciers also covered

large regions of Canada Europe, and Asia, and they are responsible for many distinctive topographic features in these regions (Section 17.2 and 17.3).

Continent glaciers can form and grow when climate conditions in a region cool over extended periods of time. Snow can build up over time in regions that do not warm up seasonally, and if the snow accumulates in vast amounts, it can compact under its own weight and form ice.

Earth's two current continental glaciers, the Antarctic and Greenland Ice Sheets, comprise about 99% of Earth's glacial ice, and approximately 68% of Earth's fresh water. The Antarctic Ice Sheet is vastly larger than the Greenland Ice Sheet (Figure 17.4) and contains about 17 times as much ice. If the entire Antarctic Ice Sheet melted, sea level would rise by about 80 m and most of Earth's major coastal cities would be submerged.

Figure 17.4 Simplified cross-section profiles of the Antarctic and Greenland continental ice sheets. Both ice sheets are drawn to the same scale (exaggerated in the vertical direction). *Source: Steven Earle (2015) CC BY 4.0. https://opentextbc.ca/geology/*

Continental glaciers generally cover areas that are flat, but the force of gravity still acts on them and causes them to flow. Continental glacier ice flows from the region where it is thickest toward the edges where it is thinner (Figure 17.5). In the central thickest parts, the ice flows almost vertically down toward the base, while at the edges of the glacier, it flows horizontally out toward the margins. In continental glaciers like the Antarctic and Greenland Ice Sheets, the thickest parts (4,000 m and 3,000 m thick, respectively) are the areas where the rate of snowfall, and therefore of ice accumulation, are greatest. In Antarctica, the ice sheet flows out over the ocean, forming ice shelves. Ice shelves can slow the flow of continental glaciers outward. Conversely, if ice shelves break down continental glacier flow can speed up.

Figure 17.5 Cross-section showing ice-flow in the Antarctic Ice Sheet. *Source: Steven Earle (2015) CC BY 4.0. https://opentextbc.ca/geology/*

17.1.2 Alpine Glaciers

Alpine glaciers (aka valley glaciers) originate high up in the mountains, mostly in temperate and polar regions (Figure 17.1), but also in tropical regions in high mountains (e.g. in the Andes Mountains of South America).

The flow of alpine glaciers is driven by gravity, and primarily controlled by the slope of the ice surface (Figure 17.6). Alpine glaciers grow due to accumulation of snow over time. In the **zone of accumulation**, the rate of snowfall is greater than the rate of melting. In other words, not all of the snow that falls each winter melts during the following summer, and the ice surface in the zone of accumulation does not lose its annual accumulation of snow cover over the course of the year. In the **zone of ablation**, the rate of melting exceeds accumulation. The **equilibrium line** marks the boundary between the zones of accumulation (above) and ablation (below) (Figure 17.6).

Figure 17.6 Schematic diagram illustrating alpine glacier ice-flow. *Source: Steven Earle (2015) CC BY 4.0. https://opentextbc.ca/geology/*

Above the equilibrium line of a glacier, winter snow will remain even after summer melting, so snow gradually accumulates on the glacier over time. The snow layer from each year is covered and compacted by subsequent snow, and it is gradually compressed and converted to **firn** (Figure 17.7). Firn is a form of ice that forms when snowflakes lose their delicate shapes and become granules due to

Figure 17.7 Steps in the process of formation of glacial ice from snow, granules, and firn. *Source: Steven Earle (2015) CC BY 4.0. https://opentextbc.ca/geology/*

compression. With more compression, the granules are squeezed together, and air is forced out. Eventually the granules are "welded" together to create glacial ice (Figure 17.7). Downward percolation and freezing of water from melting contributes to the process of ice formation.

The equilibrium line of a glacier near Whistler, BC, is shown in Figure 17.8. Below this line is the zone of ablation. In the zone of ablation, bare ice is exposed because the previous winter's snow has all melted. Above this line the ice is still mostly covered with snow from the previous winter.

The position of the equilibrium line changes from year to year as a function of the balance between snow accumulation in the winter, and snow and ice melt during the summer. If there is more winter snow and less summer melting, this favours the advance of the equilibrium line down the glacier (and ultimately increases the size of the glacier). Between accumulation and melting, the summer melt matters most to a glacier's ice budget. Cool summers promote an increase in glacier size, and thus lead to advance of the equilibrium line. Warm summers promote melting, and retreat of the equilibrium line.

Alpine glaciers move because they are heavy, and the force of gravity acts on the ice in the glacier to pull it down the slope of the mountains where they form. The movement of the glacier generates stress in the ice, which is proportional to the slope of the glaciers surface features of the underlying rock surface, and to the depth within the glacier.

As shown in Figure 17.9, the stresses are relatively small near the ice surface but much larger at depth. Stresses are greater in areas where the ice surface is relatively steep.

Figure 17.8 The approximate location of the equilibrium line (red) in September 2013 on the Overlord Glacier, near Whistler, B.C. *Source: Steven Earle (2015) CC BY 4.0, after Isaac Earle CC BY 4.0. https://opentextbc.ca/geology/*

Like rock, ice behaves in a brittle fashion under low pressure conditions (shallow depths in the glacier), and plastically at higher pressures (deeper in the glacier). Stress also affects how ice deforms; at high stress ice will either break or deform plastically (ductile deformation) depending on the pressure conditions. Under brittle deformation conditions (low pressures, shallow depths in the glacier), stress is released when the ice cracks, so does not build up to high values. Within the upper 50 - 100 m of ice (above the dashed red line, in Figure 17.9), flow is brittle: the ice is rigid and will crack in response to stress. Under ductile deformation conditions (higher pressures deeper in the glacier), stress can accumulate, and the ice will flow plastically in response to that stress. Ice deforms plastically if deeper than about 100 m in the glacier, and in this region stress levels can accumulate to high values (100 kilopascals or greater, Figure 17.9).

When the lower ice of a glacier flows, it moves the upper ice along with it. It may seem from the stress patterns (red numbers and arrows in Figure 17.9) that the lower ice moves more or faster than the upper ice, but this is not the case. The lower ice deforms (flows) and the upper part is carried

along and deforms through brittle deformation if subjected to sufficient stress. The upper part of the glacier moves faster than the base of the glacier because there is friction between the base of the glacier and the surface beneath it that slows the movement of the ice at the base.

Figure 17.9 Stress within an alpine glacier (red numbers) as determined from the slope of the ice surface and the depth within the ice. The ice will deform and flow where the stress is greater than about 100 kilopascals, and regions with higher rates of deformation are depicted by the red arrows. Any motion of the lower ice will be transmitted to the ice above it, so although the red arrows get shorter toward the top, the ice is still moving (blue arrows in centre of diagram inset illustrate rate of ice motion). The upper ice (above the red dashed line) does not flow plastically, but it is carried along with the lower ice. *Source: Steven Earle (2015) CC BY 4.0, with minor modifications by Joyce McBeth*.

The plastic lower ice of a glacier can flow over irregularities in the rocks under the glacier. However, the upper rigid ice cannot flow in this way, and because it is being carried along by the lower ice, it tends to crack in locations when the lower ice flows over changes in the topography below the glacier. This leads to formation of **crevasses** in areas where the rate of flow of the deeper, plastic ice is changing. In the area shown in Figure 17.10, for example, the glacier is accelerating over the steep terrain, and the rigid surface ice cracks to release stress that accumulates due to the change in velocity and tension in the ice.

In addition to deformation, another important aspect of glacier flow is **basal sliding**, which is sliding movement between the base of the glacier and the underlying material. The base of a glacier can be cold (below the freezing point of water) or warm (above the freezing point). If it is warm, a film of water can form between the ice and the material underneath, and the ice will be able to slide over this surface (Figure 17.11, left). If the base is cold, the ice will be frozen to the material underneath and it will be stuck — unable to slide along its base. In this case, all the movement of the ice will be by internal flow.

There are several factors that can influence warming of the ice and basal flow at the base of an alpine glacier. Friction between the base of the glacier and the surface underneath generates heat and can lead to melting of the ice at the base of the glacier. Rainwater and meltwater from upper regions of the glacier can percolate down and transfer heat to warm the base of the glacier and enhance basal sliding, particularly in warmer seasons. Geothermal heat from below also contributes to melting at the base of glaciers in regions with high heat flow due to volcanic activity.

Figure 17.10 Crevasses in a glacier in Mount Cook National Park, New Zealand. *Source: Bernard Spragg, CC0 1.0. https://flic.kr/p/dn5m8C*

Another factor that controls the temperature at the base of a glacier is the thickness of the ice. The force of gravity acting on thicker ice can enhance friction and melting at the base. Ice is also a good insulator so can prevent accumulated heat from escaping. The leading edge of an alpine glacier is typically relatively thin (see Figure 17.9), so it is common for this part to be frozen to its base while the rest of the glacier is still sliding. Since the leading edge of the glacier is frozen to the ground, and the rest of the glacier behind continues to slide forward, this causes the trailing ice to be pushed (or thrust) over top of the leading edge, forming thrust faults in the ice (Figure 17.12).

Figure 17.11 Differences in glacial ice motion with basal sliding (left) and without basal sliding (right). The dashed red line indicates the upper limit of plastic internal flow. *Source: Steven Earle (2016) CC BY 4.0. https://opentextbc.ca/geology/*

Figure 17.12 Thrust faults at the leading edge of the Byron Glacier, Portage Lake, Alaska, USA. The dark stripes are sediments that were entrained in the base of the glacier ice and transported up along the thrust faults. *Source: Cindy Zackowitz (2011) CC BY-NC 2.0. https://flic.kr/p/akCttY*

Just as the base of a glacier moves slower than the surface, the edges, which are more affected by friction along the channel walls, also move slower. If we were to place a series of markers across an alpine glacier and come back a year later, we would see that the ones in the middle had moved further forward than the ones near the edges (Figure 17.13).

Figure 17.13 Markers on an alpine glacier move forward at different rates over a period of time. *Source: Steven Earle (2015) CC BY 4.0. https://opentextbc.ca/geology/*

Alpine glacial ice continuously moves down the slope of the ice in response to gravity, but it may not appear to be moving because the front edge of a glacier is also continuously losing volume. It either melts or, if they glacier terminates at a lake or ocean, the front edge will **calve** into the water (break off pieces of the front edge of the glacier that become icebergs). If the rate of forward motion of the glacier is faster than the rate of **ablation** (melting), the leading edge of the glacier advances (moves forward). If the rate of forward motion is about the same as the rate of ablation, the leading edge remains stationary, and if the rate of forward motion is slower than the rate of ablation, the leading-edge retreats (moves backward).

Calving of icebergs is an important process for glaciers that terminate in lakes or oceans. An example of such a glacier is the Berg Glacier on Mt. Robson (Figure 17.14), which sheds small icebergs into Berg Lake. The Berg Glacier also lose mass by melting, evaporation, and sublimation.

Figure 17.14 Mt. Robson, the tallest peak in the Canadian Rockies, hosts the Berg Glacier (centre), and Berg Lake. Although there were no icebergs visible when this photo was taken, the Berg Glacier loses mass by shedding icebergs into Berg Lake. *Source: Steven Earle (2015) CC BY 4.0. https://opentextbc.ca/geology/*

Exercise 17.1 Ice Advance and Retreat

These diagrams represent a glacier with markers placed on its surface to determine the rate of ice motion over a one-year period. The ice is flowing from left to right.

1. In the middle diagram, the leading edge of the glacier has advanced. Draw in the current position of the markers.

2. In the lower diagram, the leading edge of the glacier has retreated. Draw in the current position of the markers.

17.2 Glacial Erosion

Glaciers are effective agents of erosion, especially in situations where the base of the glacier is not frozen to the underlying material and can therefore slide over the bedrock or other sediment. The ice itself is not particularly effective at erosion because it is relatively soft (Mohs hardness 1.5 at 0°C). Glacial erosion is primarily driven by abrasion of the underlying rocks by rock fragments embedded within the ice. These rocks are pushed down onto the underlying surfaces by the ice, and because they are hard they can gouge and grind down the materials beneath the glacier. An analogy for these processes is to compare the effect of a regular piece of paper being rubbed against a wooden surface ("ice eroding rock") to rubbing a piece of sandpaper over the same surface ("ice with embedded rocks

eroding rock"). The results of glacial erosion are different in areas with continental glaciation versus alpine glaciation.

17.2.1 Continental Glacial Erosion Features

Continental glaciation tends to produce relatively flat bedrock surfaces, especially where the rock beneath is uniform in strength. In areas where there are differences in the strength of rocks, a glacier tends to erode the softer and weaker rock more effectively than the harder and stronger rock. Much of central and eastern Canada, which was completely covered by the huge Laurentide Ice Sheet at various times during the Pleistocene Epoch, has been eroded to a relatively flat surface. Glacial deposits have created distinctive topographic features on the landscapes in these regions — such as drumlins, eskers, and moraines (Figure 17.16). These continental glacial features are deposits of glacial materials and are described further in Section 17.3.

Figure 17.16 Landscape features associated with continental glaciation Source: Luis María Benítez (2005) CC-BY 4.0. *https://commons.wikimedia.org/wiki/File:Receding_glacier_landscape_LMB.png*

In areas of continental glaciation, the lithosphere is depressed by the weight of glacial ice that is up to 4,000 m thick. Basins formed along the edges of continental glaciers; for example, basins formed around the edges of the Laurentide Ice Sheet that once covered much of Canada (Section 17.4). These basins filled with glacial meltwater, and layers of sediments. Many such lakes, some of them huge, existed at various times along the southern edge of the Laurentide Ice Sheet.

One example of these lakes was Glacial Lake Missoula, which formed within Idaho and Montana, just south of the BC border with the United States. During the latter part of the last glaciation (30 ka to 15 ka), the ice holding back Lake Missoula retreated enough to allow some of the lake water to escape, which escalated into a voluminous and rapid outflow (over days to weeks). During this outflow, most of the lake drained into the Columbia River valley and flowed to the Pacific Ocean. It is estimated that this type of catastrophic outflow happened at least 25 times during this period, and in many cases, the rate of outflow was equivalent to the discharge of all of Earth's current rivers combined.

17.2.2 Alpine Glacial Erosion Features

Alpine glaciers produce very different topography than continental glaciers. Alpine glaciers produce wide valleys with relatively flat bottoms and steep sides due to the erosion that occurs at the base and edges of the glaciers. These are known as **U-shaped valleys** (Figure 17.17). In contrast, unglaciated river valleys generally have a V shape.

In coastal regions where the bottom of the valley is filled with water, the U-shaped valleys are called **fjords**. The coastal mountains of BC have many fine examples of U-shaped valleys and fjords. For example, Howe Sound is a fjord that was once occupied by a large glacier. Howe Sound and most of its tributary valleys have pronounced U-shaped profiles due to glaciation (Figure 17. 18).

Figure 17.17 (Above) A depiction of a U-shaped valley occupied by a large glacier. *Source: Steven Earle (2016) CC BY 4.0. https://opentextbc.ca/geology/*

Figure 17.18 (Right) The view down the U-shaped valley of Mill Creek valley toward the U-shaped valley of Howe Sound, with the village of Britannia on the opposite side. *Source: Keefer4 (2005) CC BY-SA 2.5. http://commons.wikimedia.org/wiki/ File:Woodf1a.jpg*

Several other topographic features derived from alpine glacial erosion are found in U-shaped valleys and their tributary valleys (Figure 17.19). **Arêtes** are sharp ridges formed between U-shaped glacial valleys. **Cols** are low points (saddles) along arêtes; they form passes (high points) between glacial valleys**. Horns** are steep peaks that have been eroded by glaciers and

Figure 17.19 A diagram of some of the important alpine-glaciation erosion features. *Source: Steven Earle (2015) after Luis María Benítez, CC0 1.0. http://commons.wikimedia.org/wiki/ File:Glacial_landscape_LMB.png*

freeze-thaw activity on three or more sides. **Cirques** are bowl-shaped basins that form at the head of a glacial valley, and **tarn**s are lakes that form when cirques are flooded. **Hanging valleys** form when U-shaped valleys of tributary glaciers connect with a larger U-shaped valley; the tributary valley hangs above the main valley because the larger main-valley glacier is eroded more deeply into the terrain. **Truncated spurs** (aka "spurs") are features at the ends of arêtes where the rock is eroded into steep triangle-shaped cliffs by the glacier in the main valley.

Figure 17.20 shows examples of these features in the Swiss Alps. The area in the image was intensely glaciated during the past glacial maximum and still contains glaciers. The large U-shaped valley in the lower right was occupied by glacial ice historically, and all of the other glaciers shown here were longer and much thicker than they are now. But even at the peak of the Pleistocene glaciation, some of the higher peaks and ridges in this image would have been exposed and not directly affected by glacial erosion. A peak that extends above the surrounding glacier is called a **nunatak**. In these areas, and in the areas above the glaciers in the image today, most of the erosion is linked to freeze-thaw action.

A **roche moutonnée** is aglacial erosion feature that forms when a glacier moves over an outcrop of bedrock. Roche moutonnées consist of a hill of rock, often with a smooth, often low angle slope on one side, and a steeper and jagged slope on the other side. The side that is smooth and relatively low angle is the side the glacier was flowing from, and the a steep and sometimes jagged side is the direction the ice was moving (Figure 17.21, left). **Glacial grooves** (tens of centimetres to metres wide) and **glacial striae** (millimetres to centimetres wide) are created by the erosion caused by fragments of rock embedded in the ice at the base of a glacier (Figure 17.22, left and right). Glacial striae are very common on rock surfaces eroded by both alpine and continental glaciers. **Glacial polish** occurs when the abrasion of the rock by the glacier renders the rock so smooth it reflects light.

Figure 17.20 A view of the Swiss Alps from the International Space Station taken in 2006. The region shown is in the area of the Aletsch Glacier. The prominent peaks labelled "Horn" are the famous mountain peaks the Eiger (left) and Wetterhorn (right). A variety of alpine glacial erosion features are labelled. *Source: Steven Earle (2016) after NASA Earth Observatory (n.d.) Public Domain. http://earthobservatory.nasa.gov/IOTD/view.php?id=7195*

Figure 17.21 Roche moutonnée near Myot Hill, Scotland. *Source: Chris Upson (2006) CC BY-SA 2.0. https://en.wikipedia.org/wiki/R%C3%B4che_moutonn%C3%A9e*|

Figure 17.22 Examples of glacial striae from near Squamish, BC. Ice flow was from right to left in both images. *Source: Steven Earle (2016) CC BY 4.0. https://opentextbc.ca/geology/*

Figure 17.23 Lower Thornton Lake, a tarn, in the Northern Cascades National Park, Washington. *Source: Jeff Pang (2007) CC BY 2.0. https://www.flickr.com/photos/jeffpang/3175428205/*

Figure 17.24 Peyto Lake in the Alberta Rockies, is both a finger lake and a moraine lake, as it is flooding a glacial valley, and is dammed by an end moraine at right. *Source: Jeff Hollet (2016) Public Domain https://flic.kr/p/HJVqwo*

Lakes are common features in glacial environments. A lake that is confined to a glacial cirque is known as a **tarn** (Figure 17.23). Tarns are common in areas of alpine glaciation because the ice that forms a cirque typically carves out a depression in bedrock that can then fill with water. Moraines, which are linear deposits of glacial sediments (till) left by the glacier along its edges, can form a dam at the end of a tarn. Often, a series of moraines will form as glaciers recede. These can act as water dams, and result in strings of lakes called **rock basin lakes** or **paternoster lakes**.

A lake that occupies a glacial valley is known as a **finger lake**. In some cases, a finger lake is confined by a dam formed by an end moraine, in which case it may be called a **moraine lake** (Figure 17.24). Another type of glacial lake is a **kettle lake**. These are discussed in section 17.4 in the context of glacial deposits.

Exercise 17.2 Identify Glacial Erosion Features

This is a photo of Mt. Assiniboine in the BC Rocky Mountains. What are the features at locations **a** through **e**? Look for one of each of the following: a horn, an arête, a truncated spur, a cirque, and a col. Try to identify some of the numerous other arêtes in this view, as well as another horn.

Figure 17.25 Glacial erosion features of Mt. Assiniboine. Source: Steven Earle (2015) after Kurt Stegmüller (2008) CC BY 3.0. *https://commons.wikimedia.org/wiki/File:Mount_Assiniboine_Sunburst_Lake.jpg https://opentextbc.ca/geology/*

17.3 Glacial Deposits

Sediments transported and deposited during glaciations are abundant throughout Canada. They are important sources of aggregate for construction materials (sand, gravel), and are also important groundwater reservoirs. Because they are almost all unconsolidated, they have significant implications for slope stability and mass wasting. Figure 17.26 illustrates some of the ways that sediments are transported and deposited by alpine glaciers. The Bering Glacier is the largest glacier in North America, and although most of it is in Alaska, it flows from an icefield that extends into the southwestern Yukon Territory. The surface of the ice is partially, or in some cases completely, covered with rocky debris that has fallen onto the glacier from surrounding steep rock faces. There are muddy rivers issuing from the glacier in several locations, depositing sediment on land, into Vitus Lake, and directly into the ocean. Icebergs are portions of the glacier that have broken off and float away in a lake or ocean. Icebergs are laden with glacial sediments, which are released and deposited as the icebergs melt. Also, not visible in this view, there are sediments being moved along within and beneath the glacier itself.

Sediments are formed and transported in several ways in glacial environments (Figure 17.27). There are many different kinds of glacial sediments, which are generally classified by whether they are transported on, within, or beneath the glacial ice.

Figure 17.26 Part of the Bering Glacier in southeast Alaska, the largest glacier in North America. It is about 14 km in width in the centre of this view. *Source: NASA Earth Observatory, 2002. https://earthobservatory.nasa.gov/IOTD/view.php?id=4710*

Supraglacial (on top of the ice) and **englacial** (within the ice) sediments are released from the melting front of a stationary glacier. These sediments can form a ridge of unsorted sediments called an **end moraine**. The end moraine from the furthest advance of a glacier is called a **terminal moraine**. The general name for any sediments transported and deposited by glacial ice is **till.**

Subglacial sediment (e.g., **lodgement till**) is material that has been eroded from the rock underlying the glacier by the ice and then transported by the ice. It has a wide range of grain sizes, including a relatively high proportion of silt and clay. The larger clasts (pebbles to boulders in size) tend to become partly rounded by abrasion. When a glacier eventually melts, the lodgement till is exposed as a sheet of well-compacted sediment ranging from several centimetres to many metres in thickness. Lodgement till is normally poorly sorted and does not contain bedding features like a lake or stream sediment (Figure 17.28).

Supraglacial sediments are primarily derived from freeze-thaw eroded material that has fallen onto the ice from rocky slopes above. These sediments form **lateral moraines** (moraine deposits along the

edges of the glacier, see Figure 17.1 for an example). Where two glaciers meet, the sediments form **medial moraines** (medial moraines are visible in Figure 17.20 and Figure 17.26.) Most of this material is deposited on the ground when the ice melts. This is called **ablation till**, a mixture of fine and coarse angular rock fragments, with much less sand, silt, and clay than lodgement till (Figure 17.28). When supraglacial sediments become incorporated into the body of the glacier, they are known as englacial sediments (Figure 17.27).

Figure 17.27 A depiction of the various types of sediments associated with the Bering Glacier. The glacier is shown in cross-section. *Source: Steven Earle (2016) CC BY 4.0 https://opentextbc.ca/geology/*

Water flows on the surface, within, and at the base of a glacier, even in cold areas and even when the glacier is advancing. Depending upon its velocity, this water is able to transport sediments of various sizes, and discharges most of these sediments out of the lower end of the glacier, where they are deposited as **outwash sediments**. These sediments accumulate in a wide range of environments in the **proglacial** region (the area in front of a glacier). Most of the sediments accumulate in fluvial environments, but some are deposited in lacustrine and marine environments. **Glaciofluvial sediments** are similar to sediments deposited in normal fluvial environments, but are glacially-derived sediments, and are thus dominated by silt, sand, and gravel. The grains tend to be moderately well rounded and sorted, and the sediments have similar sedimentary structures (e.g., bedding, cross-bedding, clast imbrication [overlapping]) to those formed by non-glacial streams (Figure 17.29).

A large proglacial plain of sediment is called a **sandur** (aka **outwash plain**), and within this area, glaciofluvial deposits can be tens of metres thick. In situations where a glacier is receding, a block of ice might become separated from the main ice sheet and become buried in glaciofluvial sediments. When the ice block eventually melts, a depression forms, known as a **kettle**, and if this fills with water, it is known as a **kettle lake** (Figure 17.30, 17.32). Kettle lakes are also known as pothole lakes or prairie potholes.

Figure 17.28 Examples of glacial till: a: lodgement till from the front of the Athabasca Glacier, Alberta; b: ablation till at the Horstman Glacier, Blackcomb Mountain, BC. *Source: Steven Earle (2015) CC BY 4.0. https://opentextbc.ca/geology/*

Figure 17.29 Examples of glaciofluvial sediments: a: glaciofluvial cross-bedded sand of the Quadra Sand Formation at Comox, BC.; b: glaciofluvial gravel and sand, Nanaimo, BC. *Source: Steven Earle (2016) CC BY 4.0. https://opentextbc.ca/geology/*

A supraglacial, englacial, or subglacial stream will create its own channel within the ice, and sediments that are being transported and deposited by the stream will build up within that channel. When the ice melts, the sediment will be deposited upon the underlying ground surface to form a long sinuous ridge known as an **esker**. Eskers are most common in areas of continental glaciation. They can be several metres high, tens of metres wide, and tens of kilometres long (Figure 17.31). Eskers are commonly comprised of well-sorted sands.

Figure 17.30 A kettle lake amid vineyards and orchards in the Osoyoos area of BC. *Source: Steven Earle (2015) CC BY 4.0.* https://opentextbc.ca/geology/

Figure 17.31 Part of an esker that formed beneath the Laurentide Ice Sheet in northern Canada. *Source: Gord McKenna (1986) CC BY-NC-ND 2.0. https://flic.kr/p/9WTG5j*

Figure 17.32 (Below) Drumlins and kettle lakes viewed from the air near Fort St John, BC. There are numerous drumlins in the image; one is outlined in red. Can you spot the others? Note the alignment of the long axes of the drumlins. *Source: Joyce McBeth (2002) CC-BY 4.0.*

Drumlins are elongated, oval shaped ridges of englacial to subglacial sediments that form at the base of continental glaciers. They are often tens of metres high and hundreds of metres long, and often

occur in clusters ("fields") of tens to hundreds of drumlins (Figure 17.32). As the sediments are deposited, the glacier molds the drumlins' shapes as the glacier moves over and around them. The long axis of a drumlin is aligned with the direction that the ice moved when the drumlin was deposited.

Glacial outwash streams commonly flow into **proglacial lakes** (lakes in front of glaciers) where

Figure 17.33 Examples of glaciolacustrine and glaciomarine sedimentary structures. a: varved glaciolacustrine sediments containing a drop stone, Nanaimo, BC.; and b: a laminated glaciomarine sediment, Englishman River, BC. *Source: Steven Earle (2015) CC BY 4.0. https://opentextbc.ca/geology/*

glaciolacustrine sediments are deposited. These are dominated by silt- and clay-sized particles and are typically laminated (finely layered) on the millimetre scale. In some cases, **varves** develop. Varves are a series of beds with distinctive summer and winter layers: relatively coarse in the summer when melt discharge is high, and finer in the winter, when discharge is low. Icebergs are common in proglacial lakes, and most of them contain englacial sediments of various sizes. As the icebergs melt, the released clasts sink to the bottom and are incorporated into the glaciolacustrine layers as **drop stones** (Figure 17.33a). The processes that occur in proglacial lakes can also take place where a glacier terminates in the ocean. The sediments deposited there are called **glaciomarine sediment**s (Figure 17.33b).

Exercise 17.3 Identify Glacial Depositional Environments

Refer to the photo of the Bering Glacier in Alaska shown in Figure 17.26. Glacial sediments of many different types are being deposited throughout the region depicted in this photo.

Identify where you would expect to fine the following:
(a) glaciofluvial sand
(b) lodgement till
(c) glaciolacustrine clay with drop stones
(d) ablation till
(e) glaciomarine silt and clay

17.4 Glaciations over Earth's History

We are currently living in the middle of a **glacial period**, though it is less intense now than it was 20,000 years ago. This is not the only period of glaciation in Earth's history; there have been many in the distant past (Figure 17.34). In general, however, over the course of Earth's history the Earth's surface has been warm and ice-free for longer periods than it has been cold and glaciated.

Figure 17.34 The record of major past glaciations during Earth's history. Source: *Steven Earle (2015) CC BY 4.0. https://opentextbc.ca/geology/*

17.4.1 Pre-Cenozoic Glaciations

The oldest known glacial period is the Huronian. Based on evidence of glacial deposits from the area around Lake Huron in Ontario and elsewhere, it is evident that the Huronian Glaciation lasted from approximately 2.4 to 2.1 Ga. Because rocks of that age are rare, we do not know much about the intensity or global extent of this glaciation.

Late in the Proterozoic, for reasons that are not fully understood, the climate cooled dramatically, and Earth had the most intense time of glaciation it has ever experienced. The glaciations of the Cryogenian Period (*cryo* is Latin for icy cold) are also known as the "Snowball Earth" glaciations. Scientists have hypothesized that the entire planet was frozen at this time — even in equatorial regions — with ice on the oceans up to 1 km thick. A visitor to our planet at that time would not have found it habitable, although life still survived in the oceans.

There were two main glacial periods within the Cryogenian, each lasting for about 20 million years: the Sturtian at around 700 Ma and the Marinoan at 650 Ma. There is also evidence of some shorter glaciations both before and after these longer periods of glaciation. The end of the Cryogenian glaciations coincides with the evolution of relatively large and complex life forms on Earth. This started during the Ediacaran Period, and then continued with the so-called explosion of life forms in the Cambrian. Some geologists think that the changing environmental conditions of the Cryogenian are what triggered the evolution of large and complex life.

There have been three major glaciations during the Phanerozoic (the past 540 Ma). These include the Andean/Saharan (recorded in rocks of South America and Africa), the Karoo (named for rocks in southern Africa), and the Cenozoic glaciations. The Karoo was the longest of the Phanerozoic glaciations, persisting for much of the time that the supercontinent Gondwana was situated over the

South Pole (~360 to 260 Ma). Glaciers covered large parts of Africa, South America, Australia, and Antarctica. This widespread glaciation, across continents that are now far apart, was an important component of Alfred Wegener's evidence for continental drift. Unlike the Cryogenian glaciations, the Andean/Saharan, Karoo, and Cenozoic glaciations only affected parts of Earth. During Karoo times, for example, what is now North America was near the equator and remained unglaciated.

Earth was warm and essentially unglaciated throughout the Mesozoic. Although there may have been some alpine glaciers at this time, there is no evidence for them preserved in the geologic record. The dinosaurs, which dominated terrestrial habitats during the Mesozoic, did not have to endure icy conditions.

17.4.2 Cenozoic Glaciations

A warm climate persisted into much of the Cenozoic; there is evidence that the Paleocene (from about 50 to 60 Ma) was the warmest part of the Phanerozoic since the Cambrian (Figure 17.35). A number of tectonic events during the Cenozoic have contributed to persistent and significant planetary cooling from 50 Ma to near the present. For example, the collision of the Indian plate with the Eurasian plate and the formation of the Himalayan range and the Tibetan Plateau. Mountain building events such as the formation of the Himalayas are followed by weathering and erosion of the uplifted rocks. Higher than normal global rates of silicate mineral weathering associated with mountain building, especially weathering of feldspar, leads to a decrease in carbon dioxide concentrations in the atmosphere. This contributes to global climate cooling.

Figure 17.35 Global temperature trends over the past 65 Ma (the Cenozoic). From the end of the Paleocene to the height of the Pleistocene glaciation, global average temperature dropped by about 14°C. *Source: Joyce McBeth (2018) CC BY 4.0. Modified after Steven Earle (2015) and Makiko Sato & James Hansen (*2012), *including data from Zachos et al (2008) http://www.columbia.edu/~mhs119/Sensitivity+SL+CO2/ https://opentextbc.ca/geology/*

At 40 Ma, ongoing plate motion widened the narrow gap between South America and Antarctica, resulting in the opening of the Drake Passage. This allowed for unrestricted west-to-east flow of water around Antarctica: the Antarctic Circumpolar Current (Figure 17.36), which effectively isolated the

Southern Ocean from the warmer waters of the Pacific, Atlantic, and Indian Oceans. The region cooled significantly, and by 35 Ma (Oligocene) glaciers had started to form on Antarctica.

Global temperatures remained relatively steady during the Oligocene and early Miocene, and the Antarctic glaciation waned during that time. At around 15 Ma, subduction-related volcanism between central and South America created the land connection between North and South America, preventing water from flowing between the Pacific and Atlantic Oceans. This further restricted ocean currents that transfer heat from the tropics to the poles, leading to cooling and advance of the Antarctic glaciation. The expansion of the Antarctic ice sheet increased reflection of solar radiation at the Earth's surface and promoted a positive feedback loop of further cooling. With more reflective glacial ice, there was more cooling, leading to accumulation of more ice, and so on. By the Pliocene (~5 Ma) ice sheets had started to grow in North America and northern Europe. The most intense part of the current glaciation — and the coldest climate conditions of the current glaciation — has been during the past million years (the last third of the Pleistocene).

Figure 17.36 The Antarctic Circumpolar Current (red arrows) prevents warm water from the rest of Earth's oceans from reaching Antarctica. *Source: Steven Earle (2015) CC BY 4.0. https://opentextbc.ca/geology/*

Figure 17.37 Foram oxygen isotope record for the past 5 million years based on O isotope data from sea-floor sediments *Source: Steven Earle (2015) CC BY 4.0 using data from Lisiecki and Raymo (2005). https://opentextbc.ca/geology/*

The Pleistocene Epoch of the Cenezoic Era (2.58 Ma to 0.126 Ma), is also known as the Ice Age, Pleistocene Glaciation, or Quaternary Glaciation. The Pleistocene has been characterized by temperature fluctuations over a range of almost 10°C on time scales of 40,000 to 100,000 years. These temperature variations have corresponding with expansion and contraction of ice sheets. The

temperature variations are attributed to subtle changes in Earth's orbit, tilt, and wobble. These cyclical changes are called Milankovitch cycles. Over the past million years, the glaciation cycles have cycled over every 100,000 years, approximately (Figure 17.37).

17.4.3 The Wisconsinan Glaciation

The **Wisconsinan Glaciation** was the last major continental glaciation in *North America* (from 150-50 ka). During the Wisconsinan, all of Canada and a small portion of the northern United States was covered with continental glaciers (Figure 17.38). The massive **Laurentide Ice Sheet** covered most of eastern Canada, as far west as the Rockies, and the smaller **Cordilleran Ice Sheet** covered most of the western region of present day BC and the Yukon Territory. At various other glacial peaks during the Pleistocene and Pliocene, the ice extent was similarly distributed over North America, and in some cases, was even more extensive. The combined Laurentide and Cordilleran Ice Sheets were comparable in volume to the current Antarctic Ice Sheet.

Figure 17.38 Extent of northern hemisphere ice sheets near the peak of the Wisconsinan Glaciation (grey shading). Interglacial ice is shown in black. *Source: Hannes Grobe (2008) CC BY 3.0 https://commons.wikimedia.org/wiki/File:Iceage_north-intergl_glac_hg.png*

Exercise 17.4 Pleistocene Glacial and Interglacial Periods

Figure 17.39 shows the past 500,000 years of the data set used in Figure 17.37. The last five glacial periods are marked with snowflakes. The most recent glaciation, which peaked at around 20 ka, is known as the Wisconsinan Glaciation. Describe the nature of temperature change that followed each of these glacial periods.

The current interglacial period (Holocene) is marked with an H. Point out the previous five interglacial periods.

Figure 17.39 Global mean temperatures over the last 500,000 years. *Source: Steven Earle (2015) CC BY 4.0, using data from http://www.lorraine-lisiecki.com/stack.html, Lisiecki and Raymo (2005). https://opentextbc.ca/geology/*

Chapter Summary

The topics covered in this chapter can be summarized as follows:

17.1 Types of Glaciers

The two main types of glaciers are continental glaciers, which are very large and cover large parts of continents (e.g. the Antarctic Ice Sheet), and alpine glaciers, which occupy mountainous regions. Ice accumulates at higher elevations — above the equilibrium line — where the snow that falls in winter does not all melt in summer. In continental glaciers, ice flows outward from where it is thickest. In alpine glaciers, ice also flows from thicker to thinner regions in the glacier, obeying the law of

gravity. At depth in glacier ice, flow occurs through internal deformation, but glaciers that have liquid water at their base can also flow by basal sliding. Crevasses form in the rigid surface ice in places where the lower plastic ice is changing flow rate or shape as it moves over the underlying topography.

17.2 Glacial Erosion

Glaciers are important agents of erosion. Continental glaciers tend to erode land surface into flat plains, while alpine glaciers create a wide variety of different erosional features. The key feature of alpine glacial erosion is the U-shaped valley. Arêtes are sharp ridges that form between two valleys, and horns form where a mountain is glacially eroded on at least three sides. Since tributary glaciers do not erode as deeply as main-valley glaciers, hanging valleys exist where the two meet. On a smaller scale, both types of glaciers form roche moutonnées, d glacial grooves, and striae.

17.3 Glacial Deposits

Glacial deposits form as materials are transported and deposited in a variety of different ways in a glacial environment. Sediments that are moved and deposited directly by ice are known as till. Till deposits left at the edges of the glacier as it recedes are known as moraines. Till can also form features such as drumlins (oval-shaped elongated hills) and kettle lakes. Glaciofluvial sediments are deposited by glacial streams, either forming eskers or large proglacial plains known as sandurs. Glaciolacustrine and glaciomarine sediments originate within glaciers and are deposited in lakes and oceans, respectively.

17.4 Glaciations over Earth's History

There have been many glaciations in Earth's past, the oldest known starting about 2.4 Ga. The late Proterozoic "Snowball Earth" glaciations were thought to be sufficiently intense to affect the entire planet. The Pleistocene Glaciation was a series of glacial events over the past 2.85 Ma. The periodicity of glaciations in the Pleistocene is related to subtle changes in Earth's orbital characteristics (Milankovitch cycles), which are exaggerated by positive climate feedback processes. North America was most recently glaciated during the Wisconsinan Glaciation, from 150-50 ka.

Review Questions

1. What feature of a continental glacier causes it to flow?
2. Explain how alpine glacier ice flows.
3. What does the equilibrium line represent in a glacier?
4. Which of the following is more important to the growth of a glacier: very cold winters or relatively cool summers? Why?
5. Describe the relative rates of ice flow, and why the rate distributions are the way they are, within the following parts of a glacier: (a) the bottom versus the top and (b) the edges versus the middle.
6. Figure 17.40 shows thrust faults in the leading edge of the Athabasca glacier in Alberta. Use a red pen to draw a dotted line along the thrust faults in the glacier. Note: they are harder to

distinguish in this image than in Figure 17.12 because they don't have sediments along the faults.

Figure 17.40 Edge of the Athabasca glacier in Alberta. *Source: Steven Earle (2015) CC BY 4.0. https://opentextbc.ca/geology/*

7. What ice conditions are necessary for basal sliding to take place?
8. What sources of heat can lead to melting and/or water accumulation at the base of a glacier?
9. Why do glaciers carve U-shaped valleys, and how does a hanging valley form? How is a river valley different than a glacial valley?
10. A horn is typically surrounded by cirques. What is the minimum number of cirques you would expect to find around a horn?
11. A drumlin and a roche moutonnée are both streamlined glacial erosion features. How do they differ in shape? How can you determine the direction of glacial flow from their shapes?
12. What are drop stones, and under what circumstances are they likely to form?
13. What types of glacial sediments are likely to be sufficiently permeable to make good aquifers?
14. Why are the Cryogenian glaciations called Snowball Earth?
15. Earth cooled dramatically from the end of the Paleocene until the Holocene. Describe some of the geological events that contributed to this cooling.
16. When and where was the first glaciation of the Cenozoic?
17. Describe the extent of the Laurentide Ice Sheet during the height of the last Pleistocene glacial period.
18. Four examples of glacial sediments are shown in Figure 17.41 below. Describe the important characteristics (e.g., sorting, layering, grain-size range, grain shape, sedimentary structures) of each, and give each a name (choose from glaciofluvial, glaciolacustrine, lodgement till, ablation till, and glaciomarine).

Figure 17.41 Examples of glacial sediments. *Source: Steven Earle (2015) CC BY 4.0. https://opentextbc.ca/geology/*

References

Hansen, J. E., and Sato, M. (2012). *Climate Sensitivity Estimated From Earth's Climate History*. Retrieved from http://www.columbia.edu/~jeh1/mailings/2012/20120508_ClimateSensitivity.pdf

Lisiecki, L. E., and M. E. Raymo (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic $d^{18}O$ records. *Paleoceanography*, 20, PA1003. doi:10.1029/2004PA001071.

Zachos, J. C., Dickens, G. R., and Zeebe, R. E. (2008). An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature 541*, 279-283. doi:10.1038/nature06588

Physical Geology, 2nd Adapted Edition by Karla Panchuk is used under a CC-BY-ND 4.0 International license.

Chapter 18. Geological Resources

Introduction

Learning Objectives

After carefully reading this chapter, completing the exercises within it, and answering the questions at the end, you should be able to:

1. Describe the importance of geological resources to our way of life
2. Summarize the types of materials mined in Canada and explain some of the processes involved in the formation of metal deposits
3. Explain how a metal deposit is developed into a mine
4. Define acid rock drainage (ARD) and discuss why some mines can lead to ARD and contamination of the environment by metals
5. Summarize some of the important industrial materials extracted in Canada and describe what they are used for
6. Describe the processes that lead to the formation of coal deposits
7. Explain the processes that lead to the formation of oil and gas, the distinction between source rocks and reservoir rocks, and the importance of traps
8. Describe the origins and recovery of some of the unconventional fossil fuels
9. Describe the origins, discovery, and extraction of diamonds in Canada

18.1 If You Can't Grow It, You Have to Mine It

It has been said that "if you can't grow it, you have to mine it," meaning that anything we can't grow we have to extract from Earth in one way or another. This includes water, of course, our most important resource, but it also includes all the other materials that we need to construct things like roads, dams, and bridges, or manufacture things like plates, toasters, and telephones. Even most of our energy resources come from Earth, including uranium and fossil fuels, and much of the infrastructure of this electrical age depends on copper (Figure 18.1).

Figure 18.1 The open pit (background) and waste-rock piles (middle) of the Highland Valley Copper Mine at Logan Lake, British Columbia [Photo by Russell Hartlaub, used with permission]

Virtually everything we use every day is made from resources from Earth. For example, let's look at a tablet computer (Figure 18.2). Most of the case is made of a plastic known as ABS, which is made from either gas or petroleum. Some tablets have a case made from aluminum. The glass of a touch screen is made mostly from quartz combined with smaller amounts of sodium oxide (Na_2O), sodium carbonate (Na_2CO_3), and calcium oxide (CaO). To make it work as a touch screen, the upper surface is coated with indium tin oxide. When you touch the screen you're actually pushing a thin layer of polycarbonate plastic (made from petroleum) against the coated glass, completing an electrical circuit. The computer is then able to figure out exactly where you touched the screen.

Computer processors are made from silica wafers (more quartz) and also include a significant amount of copper and gold. Gold is used because it is a better conductor than copper and doesn't tarnish the way silver or copper does. Most computers have nickel-metal-hydride (NiMH) batteries, which contain nickel, of course, along with cadmium, cobalt, manganese, aluminum, and the rare-earth elements lanthanum, cerium, neodymium, and praseodymium. The processor and other electronic components are secured to a circuit board, which is a thin layer of fibreglass sandwiched between copper sheets coated with small amounts of tin and lead. Various parts are put together with steel screws that are made of iron and molybdenum.

Figure 18.2 The main components of a tablet computer [SE, base photograph from https://upload.wikimedia.org/wikipedia/commons/8/8d/IPad_Air.png]

That's not everything that goes into a tablet computer, but to make just those components we need a pure-silica sand deposit, a salt mine for sodium, a rock quarry for calcium, an oil well, a gas well, an aluminum mine, an iron mine, a manganese mine, a copper-molybdenum-gold mine, a cobalt-nickel mine, a rare-earth element and indium mine, and a source of energy to transport all of the materials, process them, put them together, and finally transport the computer to your house or the store where you bought it.

Exercise 18.1 Where Does It Come From?

Look around you and find at least five objects (other than a computer or a phone) that have been made from materials that had to be mined, quarried, or extracted from an oil or gas well. Try to identify the materials involved, and think about where they might have come from. This pen is just an example.

[https://upload.wikimedia.org/wikipedia/commons/f/fd/Ballpoint-pen-parts.jpg]

18.2 Metal Deposits

Mining has always been a major part of Canada's economy. Canada has some of the largest mining districts and deposits in the world, and for the past 150 years, we have been one of the most important suppliers of metals. Extraction of Earth's resources goes back a long way in Canada. For example, the First Nations of British Columbia extracted obsidian from volcanic regions for tools and traded it up and down the coast. In the 1850s, gold was discovered in central British Columbia, and in the 1890s, even more gold was discovered in the Klondike area of Yukon. These two events were critical to the early development of British Columbia, Yukon, and Alaska.

Canada's mining sector had revenues in the order of $37 billion in 2013. The majority of that was split roughly equally among gold, iron, copper, and potash, with important but lesser amounts from nickel and diamonds (Figure 18.3). Revenues from the petroleum sector are significantly higher, at over $100 billion per year.

A metal deposit is a body of rock in which one or more metals have been concentrated to the point of being economically viable for recovery. Some **background** levels of important metals in average rocks are shown on Table 18.1, along with the typical grades necessary to make a viable deposit, and the corresponding concentration factors. Looking at copper, for example, we can see that while average rock has around 40 ppm (parts per million) of copper, a grade of around 10,000 ppm or 1% is necessary to make a viable copper deposit. In other words, copper ore has about 250 times as much copper as typical rock. For the other elements in the list, the concentration factors are much higher. For gold, it's 2,000 times and for silver it's around 10,000 times.

Figure 18.3 The value of various Canadian mining sectors in 2013 [SE from data at http://www.nrcan.gc.ca/mining-materials/publications/8772]

Table 20.1 Typical background and ore levels of some important metals [Steven Earle CC BY 4.0]

Metal	Typical Background Level	Typical Economic Grade*	Concentration Factor
Copper	40 ppm	10,000 ppm (1%)	250 times
Gold	0.003 ppm	6 ppm (0.006%)	2,000 times
Lead	10 ppm	50,000 ppm (5%	5,000 times
Molybdenum	1 ppm	1,000 ppm (0.1%)	1,000 times
Nickel	25 ppm	20,000 ppm (2%)	800 times
Silver	0.1 ppm	1,000 ppm (0.1%)	10,000 times
Uranium	2 ppm	10,000 ppm (1%)	5,000 times
Zinc	50 ppm	50,000 ppm (5%)	1,000 times

*It's important to note that the economic viability of any deposit depends on a wide range of factors including its grade, size, shape, depth below the surface, and proximity to infrastructure, the current price of the metal, the labour and environmental regulations in the area, and many other factors.

It is clear that some very significant concentration must take place to form a mineable deposit. This concentration may occur during the formation of the host rock, or after the rock forms, through a number of different types of processes. There is a very wide variety of ore-forming processes, and there are hundreds of types of mineral deposits. The origins of a few of them are described below.

Types of Metal Deposits

Magmatic Nickel Deposits

A magmatic deposit is one in which the metal concentration takes place primarily at the same time as the formation and emplacement of the magma. Most of the nickel mined in Canada comes from magmatic deposits such as those in Sudbury (Ontario), Thompson (Manitoba) (Figure 18.4), and Voisey's Bay (Labrador). The magmas from which these deposits form are of mafic or ultramafic composition (derived from the mantle), and therefore they have relatively high nickel and copper contents to begin with (as much as 100 times more than normal rocks in the case of nickel). These elements may be further concentrated within the magma as a result of the addition of sulphur from partial melting of the surrounding rocks. The heavy nickel and copper sulphide minerals are then concentrated further still by gravity segregation (i.e., crystals settling toward the bottom of the magma chamber). In some cases, there are significant concentrations of platinum-bearing minerals.

Most of these types of deposits around the world are Precambrian in age — probably because the mantle was significantly hotter at that time, and the necessary mafic and ultramafic magmas were more likely to be emplaced in the continental crust.

Figure 18.4 The nickel smelter at Thompson, Manitoba
[https://en.wikipedia.org/wiki/Thompson,_Manitoba#/media/File:Vale_Nickel_Mine.JPG]

Volcanogenic Massive Sulphide Deposits

Much of the copper, zinc, lead, silver, and gold mined in Canada comes from **volcanic-hosted massive sulphide (VHMS)** deposits associated with submarine volcanism (VMS deposits). Examples are the deposits at Kidd Creek, Ontario, Flin Flon on the Manitoba-Saskatchewan border, Britannia on Howe Sound, and Myra Falls (within Strathcona Park) on Vancouver Island.

VMS deposits are formed from the water discharged at high temperature (250° to 300°C) at ocean-floor hydrothermal vents, primarily in areas of subduction-zone volcanism. The environment is comparable to that of modern-day black smokers (Figure 18.5), which form where hot metal- and sulphide-rich water issues from the sea floor. They are called massive sulphide deposits because the sulphide minerals (including pyrite (FeS_2),

sphalerite (ZnS), chalcopyrite ($CuFeS_2$), and galena (PbS)) are generally present in very high concentrations (making up the majority of the rock in some cases). The metals and the sulphur are leached out of the sea-floor rocks by convecting groundwater driven by the volcanic heat, and then quickly precipitated where that hot water enters the cold seawater, causing it to cool suddenly and change chemically. The volcanic rock that hosts the deposits is formed in the same area and at the same general time as the accumulation of the ore minerals.

Figure 18.5 Left: A black smoker on the Juan de Fuca Ridge off the west coast of Vancouver Island. Right: A model of the formation of a volcanogenic massive sulphide deposit on the sea floor. [left: NOAA at: http://oceanexplorer.noaa.gov/okeanos/explorations/10index/background/plumes/media/black_smoker.html, right: SE]

Porphyry Deposits

Porphyry deposits are the most important source of copper and molybdenum in British Columbia, the western United States, and Central and South America. Most porphyry deposits also host some gold, which may be, in rare cases, the primary commodity. B.C. examples include several large deposits within the Highland Valley mine (Figure 18.1) and numerous other deposits scattered around the central part of the province.

A porphyry deposit forms around a cooling felsic stock in the upper part of the crust. They are called "porphyry" because upper crustal stocks are typically porphyritic in texture, the result of a two-stage cooling process. Metal enrichment results in part from convection of groundwater related to the heat of the stock, and also from metal-rich hot water expelled by the cooling magma (Figure 18.6). The host rocks, which commonly include the stock itself and the surrounding country rocks, are normally highly fractured and brecciated. During the ore-forming process, some of the original minerals in these rocks are altered to potassium feldspar, biotite, epidote, and various clay minerals. The important ore minerals include chalcopyrite ($CuFeS_2$), bornite (Cu_5FeS_4), and pyrite in copper porphyry deposits, or molybdenite (MoS_2) and pyrite in molybdenum porphyry deposits. Gold is present as minute flakes of native gold.

This type of environment (i.e., around and above an intrusive body) is also favourable for the formation of other types of deposits — particularly vein-type gold deposits (a.k.a. **epithermal deposits**). Some of the gold deposits of British Columbia (such as in the Eskay Creek area adjacent to the Alaska panhandle), and many of the other

gold deposits situated along the western edge of both South and North America are of the vein type shown in Figure 18.6, and are related to nearby magma bodies.

Figure 18.6 A model for the formation of a porphyry deposit around an upper-crustal porphyritic stock and associated vein deposits. [SE]

Banded Iron Formation

Most of the world's major iron deposits are of the **banded iron formation** type (classified as a type of chemical sedimentary rock), and most of these formed during the initial oxygenation of Earth's atmosphere between 2,400 and 1,800 Ma. At that time, iron that was present in dissolved form in the ocean (as Fe^{2+}) became oxidized to its insoluble form (Fe^{3+}) and accumulated on the sea floor, mostly as hematite interbedded with chert (Figure 18.7). Unlike many other metals, which are economically viable at grades of around 1% or even much less, iron deposits are only viable if the grades are in the order of 50% iron and if they are very large

Figure 18.7 Banded iron formation from an unknown location in North America on display at a museum in Germany. The rock is about 2 m across. The dark grey layers are magnetite and the red layers are hematite. Chert is also present. [https://upload.wikimedia.org/wikipedia/commons/5/5f/Black-band_ironstone_%28aka%29.jpg]

Unconformity-Type Uranium Deposits

There are several different types of uranium deposits, but some of the largest and richest are those within the Athabasca Basin of northern Saskatchewan. These are called **unconformity-type uranium deposits** because they are all situated very close to the unconformity between the Proterozoic Athabasca Group sandstone and the much older Archean sedimentary, volcanic, and intrusive igneous rock (Figure 18.8).

Figure 18.8 Model of the formation of unconformity-type uranium deposits of the Athabasca Basin, Saskatchewan [SE]

The origin of unconformity-type U deposits is not perfectly understood, but it is thought that two particular features are important: (1) the relative permeability of the Athabasca Group sandstone, and (2) the presence of graphitic schist within the underlying Archean rocks. The permeability of the sandstone allowed groundwater to flow through it and leach out small amounts of U, which stayed in solution in the oxidized form U^{6+}. The graphite (C) created a reducing environment (non-oxidizing) that converted the U from U^{6+} to insoluble U^{4+}, at which point it was precipitated as the mineral uraninite (UO_2).

Exercise 18.2 The Importance of Heat and Heat Engines

For a variety of reasons, thermal energy (heat) from within Earth is critical in the formation of many types of ore deposits. Look back through the deposit-type descriptions above and for each of the following, indicate whether the deposit type depends on a source of heat for their formation, and why.

a) Magmatic

b) Volcanogenic massive sulphide

c) Porphyry

d) Banded iron formation

e) Unconformity-type uranium

Mining and Mineral Processing

Metal deposits are mined in a variety of different ways depending on their depth, shape, size, and grade. Relatively large deposits that are quite close to the surface and somewhat regular in shape are mined using **open-pit mine** methods (Figure 18.1). Creating a giant hole in the ground is generally cheaper than making an underground mine, but it is also less precise, so it is necessary to mine a lot of waste rock along with the ore.

Relatively deep deposits or those with elongated or irregular shapes are typically mined from underground with deep vertical **shafts**, **declines** (sloped tunnels), and **levels** (horizontal tunnels) (Figures 18.9 and 18.10). In this way, it is possible to focus the mining on the ore body itself. However, with relatively large ore bodies, it may be necessary to leave some pillars to hold up the roof.

Above: Figure 18.9 Underground at the Myra Falls Mine, Vancouver Island. [SE]

Right: Figure 18.10 Schematic cross-section of a typical underground mine. [SE]

In many cases, the near-surface part of an ore body is mined with an open pit, while the deeper parts are mined underground (Figures 18.10 and 18.11).

Figure 18.11 Entrance to an exploratory decline (white arrow) for the New Afton Mine situated in the side of the open pit of the old Afton Mine, near Kamloops, B.C. [SE]

A typical metal deposit might contain a few percent of ore minerals (e.g., chalcopyrite or sphalerite), mixed with the minerals of the original rock (e.g., quartz or feldspar). Other sulphide minerals are commonly present within the ore, especially pyrite.

When ore is processed (typically very close to the mine), it is ground to a fine powder and the ore minerals are physically separated from the rest of the rock to make a **concentrate**. At a molybdenum mine, for example, this concentrate may be almost pure molybdenite (MoS_2). The rest of the rock is known as **tailings**. It comes out of the concentrator as a wet slurry and must be stored near the mine, in most cases, in a tailings pond.

The tailings pond at the Myra Falls Mine on Vancouver Island is shown in Figure 18.12, and the settling ponds for waste water from the concentrator are shown in Figure 18.13. The tailings are contained by an embankment. Also visible in the foreground of Figure 18.12 is a pile of waste rock, which is non-ore rock that was mined in order to access the ore. Although this waste rock contains little or no ore minerals, at many mines it contains up to a few percent pyrite. The tailings and the waste rock at most mines are an environmental liability because they contain pyrite and small amounts of ore minerals. When pyrite is exposed to oxygen and water, it generates sulphuric acid , also known as **acid rock drainage** (ARD). Acidity itself is a problem to the environment, but because the ore elements, such as copper or lead, are more soluble in acidic water than neutral water, ARD is also typically quite rich in metals, many of which are toxic.

Figure 18.12 The tailings pond at the Myra Falls Mine on Vancouver Island. The dry rock in the middle of the image is waste rock. The structure on the right is the headframe for the mine shaft. Myra Creek flows between the tailings pond and the headframe. [SE]

Figure 18.13 The tailings pond (lower left) at Myra Falls Mine with settling ponds (right) for processing water from the concentrator. [SE]

Tailings ponds and waste-rock storage piles must be carefully maintained to ensure their integrity, and monitored to ensure that acidic and metal-rich water is not leaking out. In August 2014, the tailings pond at the Mt. Polley Mine in central BC failed and 10 million cubic metres of waste water along with 4.5 million cubic

metres of tailings slurry was released into Polley Lake, Hazeltine Creek, and Quesnel Lake (Figure 18.14, a and b). As of July 2015, the environmental implications of this event are still not fully understood.

Figure 18.14a The Mt. Polley Mine area prior to the dam breach of August 2014. The tailings were stored in the area labelled "retention basin." [https://en.wikipedia.org/wiki/Mount_Polley_mine_disaster]

Figure 18.14b The Mt. Polley Mine area after the tailings dam breach of August 2014. The water and tailings released flowed into Hazeltine Creek, and Polley and Quesnel Lakes. [https://en.wikipedia.org/wiki/Mount_Polley_mine_disaster]

Most mines have concentrators on site because it is relatively simple to separate ore minerals from non-ore minerals and thus significantly reduce the costs and other implications of transportation. But separation of ore minerals is only the preliminary stage of metal refinement, for most metals the second stage involves separating the actual elements within the ore minerals. For example, the most common ore of copper is chalcopyrite ($CuFeS_2$). The copper needs to be separated from the iron and sulphur to make copper metal and that involves complicated and very energy-intensive processes that are done at **smelters** or other types of refineries. Because of their cost and the economies of scale, there are far fewer refineries than there are mines.

There are several metal refineries (including smelters) in Canada; some examples are the aluminum refinery in Kitimat, B.C. (which uses ore from overseas); the lead-zinc smelter in Trail, B.C.; the nickel smelter at Thompson, Manitoba; numerous steel smelters in Ontario, along with several other refining operations for nickel, copper, zinc, and uranium; aluminum refineries in Quebec; and a lead smelter in New Brunswick.

18.3 Industrial Minerals

Metals are critical for our technological age, but there are a lot of other not-so-shiny materials that are needed to facilitate our way of life. For everything made out of concrete or asphalt, we need sand and gravel. To make the cement that holds concrete together, we also need limestone. For the glass in our computer screens and for glass-sided buildings, we need silica sand plus sodium oxide (Na_2O), sodium carbonate (Na_2CO_3), and calcium oxide (CaO). Potassium is an essential nutrient for farming in many areas, and for a wide range of applications (e.g., ceramics and many industrial processes), we also need various types of clay.

Figure 18.15 Sand and gravel in an aggregate pit near Nanaimo, BC. [SE]

Figure 18.16 Triassic Quatsino Formation limestone being quarried on Texada Island, B.C. [SE]

The best types of **aggregate** (sand and gravel) resources are those that have been sorted by streams, and in Canada the most abundant and accessible fluvial deposits are associated with glaciation. That doesn't include till of course, because it has too much silt and clay, but it does include glaciofluvial outwash, which is present in thick deposits in many parts of the country, similar to the one shown in Figure 18.15. In a typical gravel pit, these materials are graded on-site according to size and then used in a wide range of applications from constructing huge concrete dams to filling children's sandboxes. Sand is also used to make glass, but for most types of glass, it has to be at least 95% quartz (which the sandy layers shown in Figure 18.15 are definitely not), and for high-purity glass and the silicon wafers used for electronics, the source sand has to be over 98% quartz.

Approximately 80 million tonnes of concrete are used in Canada each year — a little over 2 tonnes per person. The cement used for concrete is made from approximately 80% calcite ($CaCO_3$) and 20% clay. This mixture is heated to 1450°C to produce the required calcium silicate compounds (e.g., Ca_2SiO_4). The calcite typically comes from limestone quarries like the one on Texada Island, B.C. (Figure 18.16). Limestone is also used as the source material for many other products that require calcium compounds, including steel and glass, pulp and paper, and plaster products for construction.

Sodium is required for a wide range of industrial processes, and the most convenient source is sodium chloride (rock salt), which is mined from evaporite beds in various parts of Canada. The largest salt mine in the world is at Goderich, Ontario, where salt is recovered from the 100 m thick Silurian Salina Formation. The same formation is mined in the Windsor area. Rock salt is also used as a source of sodium and chlorine in the chemical industry to melt ice on roads, as part of the process of softening water, and as a seasoning. Under certain conditions, the mineral sylvite (KCl) accumulates in evaporite beds, and this rock is called potash. This happened across the Canadian prairies during the Devonian, creating the Prairie Evaporite Formation. Potassium

is used as a crop fertilizer, and Canada is the world's leading supplier, with most of that production coming from Saskatchewan.

Another evaporite mineral, gypsum ($CaSO_4.2H_20$), is the main component of plasterboard (drywall) that is widely used in the construction industry. One of the main mining areas for gypsum in Canada is in the Milford Station area of Nova Scotia, site of the world's largest gypsum mine.

Rocks are quarried or mined for many different uses, such as building facades (Figure 18.17), countertops, stone floors, and headstones. In most of these cases, the favoured rock types are granitic rocks, slate, and marble. Quarried rock is also used in some applications where rounded gravel isn't suitable, such as the ballast (road bed) for railways, where crushed angular rock is needed.

Figure 18.17 Slate used as a facing material on a concrete building column in Vancouver [SE]

Exercise 18.3 Sources of Important Lighter Minerals

When we think of the manufacture of consumer products, plastics and the heavy metals (copper, iron, lead, zinc) easily come to mind, but we often forget about some of the lighter metals and non-metals that are important. Consider the following elements and determine their sources. Answers for all of these except magnesium are given above. See if you can figure out a likely mineral source of magnesium.

a) Silicon
b) Calcium
c) Sodium
d) Potassium
e) Magnesium

18.4 Fossil Fuels

There are numerous types of fossil fuels, but all of them involve the storage of organic matter in sediments or sedimentary rocks. Fossil fuels are rich in carbon and almost all of that carbon ultimately originates from CO_2 taken out of the atmosphere during photosynthesis. That process, driven by solar energy, involves reduction (the opposite of oxidation) of carbon, resulting in it being combined with hydrogen instead of oxygen. The resulting organic matter is made up of complex and varied carbohydrate molecules.

Most organic matter is oxidized back to CO_2 relatively quickly (within weeks or years in most cases), but any of it that gets isolated from the oxygen of the atmosphere (for example, deep in the ocean or in a stagnant bog) may last long enough to be buried by sediments and, if so, may be preserved for tens to hundreds of millions of years. Under natural conditions, that means it will be stored until those rocks are eventually exposed at the surface and weathered.

In this section, we'll discuss the origins and extraction of the important fossils fuels, including coal, oil, and gas.

Coal

Coal was the first fossil fuel to be widely used. As we learned in section 9.4, coal forms where vigorous growth of vegetation in swampy areas leads to an abundance of organic matter that accumulates within stagnant water. The chemical nature of the water- it being acidic and having little to no oxygen- means the organic matter decays very little. If a thick layer of organic matter is accumulated and then buried, the organic matter begins to change as it is compressed and heated. This situation, where the dead organic matter is submerged in oxygen-poor water, must be maintained for centuries to millennia in order for enough material to accumulate to form a thick layer.

Over time, as the organic matter is heated and compressed more and more, the carbon within it becomes concentrated, and it can provide more energy when it is burned. Figure 18.18 shows the classification system for different grades of coal. Increasing pressure and temperature means proceeding clockwise through the diagram, starting with **lignite** (the lowest grade), then **bituminous coal**, and finally **anthracite**, the highest grade.

There are significant coal deposits in many parts of Canada, including the Maritimes, Ontario, Saskatchewan, Alberta, and British Columbia. In Alberta and Saskatchewan, much of the coal is used for electricity generation. Coal from the Highvale Mine (Figure 18.19), Canada's largest, is used to feed the Sundance and Keephills power stations west of Edmonton. Almost all of the coal mined in British Columbia is exported for use in manufacturing steel.

Figure 18.18 Coal ranking system used by the United States Geological Survey (USGS). As vegetative organic matter is buried deeper, and experiences higher pressures and temperatures, it progresses clockwise through the diagram, beginning with lignite. Additional heat and pressure result in the coal having a higher concentration of carbon (the vertical axis), and producing more energy (the horizontal axis). [USGS, public domain, https://commons.wikimedia.org/wiki/File:Coal_Rank_USGS.png#mw-jump-to-license]

Figure 18.19 The Highvale Mine (background) and the Sundance (right) and Keephills (left) generating stations on the southern shore of Wabamun Lake, Alberta [SE]

Oil and Gas

While almost all coal forms on land from terrestrial vegetation, most oil and gas is derived primarily from marine micro-organisms that accumulate within sea-floor sediments. In areas where marine productivity is high, dead organic matter is delivered to the sea floor fast enough that some of it escapes oxidation. This material accumulates in the muddy sediments, which become buried to significant depth beneath other sediments.

Figure 18.20 The depth and temperature limits for biogenic gas, oil, and thermogenic gas [SE]

As the depth of burial increases, so does the temperature in accordance with the geothermal gradient, and gradually the organic matter within the sediments is converted to hydrocarbons (Figure 18.20). The first stage is the biological production (involving anaerobic bacteria) of methane. Most of this escapes back to the surface, but some is trapped in methane hydrates near the sea floor. At depths beyond about 2 km, and at temperatures ranging from 60° to 120°C, the organic matter is converted by chemical processes to oil. This depth and temperature range is known as the **oil window**. Beyond 120°C most of the organic matter is chemically converted to methane.

The organic-matter bearing rock within which the formation of gas and oil takes place is known to petroleum geologists as the **source rock**. Both liquid oil and gaseous methane are lighter than water, so as liquids and gases form, they tend to move slowly toward the surface, out of the source rock and into **reservoir rocks**. Reservoir rocks are typically relatively permeable because that allows migration of the fluids from the source rocks, and also facilitates recovery of the oil or gas. In some cases, the liquids and gases make it all the way to the surface, where they are oxidized, and the carbon is returned to the atmosphere. In other cases, they are contained by overlying impermeable rocks (e.g., mudrock) in situations where anticlines, faults, stratigraphy changes, and reefs or salt domes create traps (Figure 18.21).

The liquids and gases that are trapped within reservoirs become separated into layers based on their density, with gas rising to the top, oil below it, and water underneath the oil. The proportions of oil and gas depend primarily on the temperature in the source rocks. Some petroleum fields, such as many of those in Alberta, are dominated by oil, while others, notably those in northeastern B.C., are dominated by gas.

Figure 18.21 Migration of oil and gas from source rocks into traps in reservoir rocks [SE]

Figure 18.22 Seismic section through the East Breaks Field in the Gulf of Mexico. The dashed red line marks the approximate boundary between deformed rocks and younger undeformed rocks. The wiggly arrows are interpreted migration paths. The total thickness of this section is approximately 5 km. [SE after http://wiki.aapg.org/File:Sedimentary-basin-analysis_fig4-55.png]

In general, petroleum fields are not visible from the surface, and their discovery involves the search for structures in the subsurface that have the potential to form traps. Seismic surveys are the most commonly used

tool for early-stage petroleum exploration, as they can reveal important information about the stratigraphy and structural geology of subsurface sedimentary rocks. An example from the Gulf of Mexico south of Texas is shown in Figure 18.22. In this area, a thick evaporite deposit ("salt") has formed domes because salt is lighter than other sediments and tends to rise slowly toward the surface; this has created traps. The sequence of deformed rocks is capped with a layer of undeformed rock

Exercise 18.4 Interpreting a Seismic Profile

The cross-section shown here is from a ship-borne seismic survey in the Bering Sea off the west coast of Alaska. As a petroleum geologist, it's your job to pick two separate locations to drill for oil or gas. Which locations would you choose?

[from USGS at: http://walrus.wr.usgs.gov/infobank/programs/html/definition/seis.html]

What Is Unconventional Oil and Gas?

The type of oil and gas reservoirs illustrated in Figures 18.21 and 18.22 are described as conventional reserves. Some **unconventional** types of oil and gas include oil sands, shale gas, and coal-bed methane.

Oil Sands

Oil sands are important in Canada because the reserves in Alberta are so large (the largest single reserve of oil in the world), but they are controversial from environmental and social perspectives. Oil sands are unconventional because the oil is exposed near the surface and is highly viscous because of microbial changes that have taken place at the surface. The hydrocarbons that form this reserve originated in deeply buried Paleozoic rocks adjacent to the Rocky Mountains and migrated up and toward the east (Figure 18.23).

The oil sands are controversial primarily because of the environmental cost of their extraction. Because the oil is so viscous, it requires heat to make it sufficiently liquid to process. This energy comes from gas; approximately 25 m^3 of gas is used to produce 0.16 m^3 (one barrel) of oil. The energy equivalent of the required gas is about 20% of the energy embodied in the produced oil. The other environmental cost of oil sands production is the

strip-mining of vast areas of land and the constructions of tailing ponds from which the release of contaminants into the groundwater and rivers of the region is unavoidable.

At present, most oil recovery from oil sands is achieved by mining the sand and processing it on site. Exploitation of oil sand that is not exposed at the surface depends on in situ processes, an example being the injection of steam into the oil-sand layer to reduce the viscosity of the oil so that it can be pumped to the surface.

Figure 18.23 Schematic cross-section of northern Alberta showing the source rocks and location of the Athabasca Oil Sands [SE]

Shale Gas

Shale gas is gas that is trapped within rock that is too impermeable for the gas to escape under normal conditions, and it can only be extracted by fracturing the reservoir rock using water and chemicals under extremely high pressure. This procedure is known as **hydraulic fracturing** or "**fracking.**" Fracking is controversial because of the volume of water used, and because, in some jurisdictions, the fracking companies are not required to disclose the nature of the chemicals used. Although fracking is typically done at significant depths, there is always the risk that overlying water-supply aquifers could be contaminated (Figure 18.24). Fracking also induces low-level seismicity (earthquakes).

Figure 18.24 Directional drilling and fracking to recover gas from impermeable rocks. Light blue arrows represent the potential for release of fracking chemicals to aquifers. [SE, after https://en.wikipedia.org/wiki/Hydraulic_fracturing#/media/File:HydroFrac2.svg

During the process that converts organic matter to coal, some methane is produced, which is stored within the pores of the coal. When coal is mined, methane is released into the mine where it can become a serious explosion hazard. Modern coal-mining machines have methane detectors on them and actually stop operating if the methane levels are dangerous. It is possible to extract the methane from coal beds without mining the coal; gas recovered this way is known as **coal-bed methane**.

18.5 Diamonds

Although Canada's diamond mining industry didn't get started until 1998, diamonds are currently the sixth most valuable product mined in the country (Figure 18.3), and Canada ranks sixth in the world in diamond production. Diamonds form deep in the mantle (approximately 200 km to 250 km depth) under very specific pressure and temperature conditions, from carbon that is naturally present in mantle rock (not from coal). The diamond-bearing rock is brought to the surface by a type of volcanism that is extremely rare (the most recent kimberlite eruption is thought to have taken place 10,000 years ago and prior to that at around 30 Ma). There is more on the volcanology of kimberlites in section 11.3. All of the world's kimberlite diamond deposits are situated within ancient shield areas (**cratons**) in Africa, Australia, Russia, South America, and North America.

It has long been known that diamonds could exist within the Canadian Shield, but up until 1991, exploration efforts had been unsuccessful. In 1980 two geologists, Chuck Fipke and Stu Blusson, started searching in the Northwest Territories by sampling glacial sediments looking for some of the minerals that are normally quite abundant within **kimberlites**: chromium-bearing garnet, chromium-bearing pyroxene, chromite (Cr_2O_3), and ilmenite ($FeTiO_3$). These distinctive minerals are used for this type of exploration because they are many times more abundant in kimberlite than diamond is. After more than a decade of exploration, Fipke and Blusson finally focused their search on an area 250 km northeast of Yellowknife, and, in 1991, they announced the discovery of a diamond-bearing kimberlite body at Lac de Gras. That discovery is now the Diavik Mine, and there is another diamond mine — Ekati — 25 km to the northwest (Figure 18.25). There are two separate mines at Diavik accessing three different kimberlite bodies, and there are five at Ekati. See Figure 11.22 for a close-up view of the Ekati Mine. There are six operating diamond mines in Canada, four in the Northwest Territories (including Diavik and Ekati), and one each in Nunavut and Ontario.

Figure 18.25 Diamond mines in the Lac de Gras region, Nunavut. The twin pits of the Diavik Mine are visible in the lower right on an island within Lac de Gras. The five pits of the Ekati mine are also visible, on the left and the upper right. The two main mine centres are 25 km apart. [http://earthobservatory.nasa.gov/IOTD/view.php?id=84085&src=eoa-iotd]

Chapter Summary

The topics covered in this chapter can be summarized as follows:

18.1 If You Can't Grow It, You Have To Mine It

Geological resources are critical to our way of life and important to the Canadian economy. Gold, iron, copper, nickel, and potash are Canada's most valuable mined commodities.

18.2 Metal Deposits

The proportions of metals in mineral deposits are typically several thousand times higher than those in average rocks, and special processes are required to extract the valuable content. Some deposits form through processes within a magma chamber, others during volcanism or adjacent to a stock, and some are related to sedimentary processes. Mining involves both surface and underground methods, but in either case, rock is brought to surface that can react with water and oxygen to produce acid rock drainage and metal contamination.

18.3 Industrial Materials

Non-metallic materials are very important to infrastructure and agriculture. Some of the major industrial minerals include sand and gravel, limestone for cement and agriculture, salt for a range of applications, potash fertilizer, and decorative stone.

18.4 Fossil Fuels

The main fossil fuels are coal, oil, and gas. Coal forms on land in wet environments where organic matter can remain submerged and isolated from oxygen for millennia before it is buried by more sediments. The depth of that burial influences the grade of coal produced. Oil and gas originate from organisms living in marine environments, and again, fairly rapid burial is required to preserve the organic matter on the sea floor. At moderate burial depth (2 km to 4 km), oil is produced, and at greater depth, gas is produced. Both oil and gas migrate toward the surface and can be trapped beneath impermeable rock layers in structural features, such as anticlines or faults. Some unconventional fossil fuel resources include oil sands, shale gas, and coal-bed methane.

18.5 Diamonds

Diamonds originate in the mantle and are only brought to the surface by the very rare eruption of kimberlitic volcanoes. The relatively recent discovery of diamonds in Canada was based on the exhaustive search for diamond indicator minerals in glacial sediments. There are now six diamond mines in Canada.

Questions for Review

1. What are some of Earth's resources that are needed to make a compact fluorescent light bulb? The image shows the contents of the ballast.

Left: https://upload.wikimedia.org/wikipedia/commons/3/31/06_Spiral_CFL_Bulb_2010-03-08_%28white_back%29.jpg Right: https://en.wikipedia.org/wiki/ Compact_fluorescent_lamp#/ media/File:Elektronstarterp.jpg

2. Explain why nickel deposits are associated only with mafic magma and not with intermediate or felsic magma?

3. What is the composition of the black smoke in a black smoker, and how does that relate to a volcanogenic massive sulphide deposit?

4. How might an epigenetic gold deposit be related to a porphyry deposit?

5. Oxidation and reduction processes are important to both banded iron formation deposits and unconformity-type uranium deposits. Explain the role in each case.

6. A typical kimberlite (right) in northern Canada may look something like the diagram shown. In this case, the diameter at the surface is around 500 m, and the total depth is about 2,500 m. Bearing in mind that an open pit cannot typically be any deeper than it is wide, what mining method(s) might be most applicable to a deposit of this type?

Kimberlite [Steven Earle CC BY 4.0]

7. What mineral is typically responsible for acid rock drainage around mine sites, and why is this mineral so common in this setting?

8. Explain why glaciofluvial gravel is more suitable than till as a source for aggregate.

9. The raw material for making cement is lime (CaO), and this is typically produced by heating limestone (mostly $CaCO_3$) to about 1,000°C. Why is this an environmental issue?

10. Name some important industrial minerals that form in an evaporite setting.

11. If organic matter accumulates at an average rate of 1 mm per year, and if 10 m of organic matter is required to make 1 m of coal, how long must a swampy environment remain stable and wet in order to form a 1.5 m coal seam?

12. What are the ideal characteristics of petroleum source rocks and petroleum reservoir rocks?

13. How deep must petroleum source rocks be buried to produce oil?

14. Why is shale gas an unconventional reserve, and how is it recovered? What are some of the environmental issues associated with that process?

15. If you were sampling glacial deposits to search for kimberlites, why would you be advised to look for kimberlite indicator minerals rather than diamonds?

Physical Geology, First University of Saskatchewan Edition is used under a CC BY-NC-SA 4.0 International License
Read this book online at http://openpress.usask.ca/physicalgeology/

Chapter 19. Measuring Geological Time

Adapted by Tim Prokopiuk and Karla Panchuk from *Physical Geology* by Steven Earle

Figure 19.1 Arizona's Grand Canyon is an icon for geological time; 1,450 million years are represented by this photo. The light-coloured layers of rocks at the top formed at ~ 250 Ma, and the dark ones at the bottom of the canyon at ~ 1,700 Ma. *Source: Steven Earle (2015) CC BY 4.0*

Learning Objectives

After reading this chapter and answering the *Review Questions* at the end, you should be able to:

- Apply basic geological principles to determine the relative ages of rocks.
- Explain the difference between relative and absolute age-dating techniques.
- Summarize the history of the geological time scale and the relationships between eons, eras, periods, and epochs.
- Understand the importance and significance of unconformities.
- Estimate the age of a rock based on the fossils that it contains.
- Use isotopic data to estimate the absolute age of a rock.
- Describe some applications and limitations of isotopic techniques for absolute geological dating.
- Describe the techniques for dating geological materials using tree rings and magnetic data.
- Explain why an understanding of geological time is critical to both geologists and the public in general.

Geological Time Is Vast

Time is the dimension that sets geology apart from most other sciences. Geological time is vast, and Earth has changed tremendously during this time. Even though most geological processes are very, very slow, the vast amount of time that has passed has allowed for the formation of extraordinary geological features, as shown in Figure 19.1.

We have numerous ways of measuring geological time. We can tell the relative ages of rocks (e.g., whether one rock is older than another) based on their spatial relationships, we can use fossils to date sedimentary rocks because we have a detailed record of the evolution of life on Earth, and we can use a range of isotopic techniques to determine the absolute ages (in millions of years) of igneous and metamorphic rocks.

But just because we can measure geological time doesn't mean that we understand it. One of the biggest hurdles faced by geology students—and geologists as well—in understanding geology is to really come to grips with the slow rates at which geological processes happen, and the vast amount of time involved.

19.1 The Geological Timescale

James Hutton (1726-1797) was a Scottish geologist, considered by some to be the father of modern Geology. Hutton studied present-day processes and applied his observations to the rock record in order to understand what he saw there. Such a method is now encapsulated in the **principle of uniformitarianism**, which states that the present is the key to the past. Given that many geological processes that we can see happening around us occur at very slow rates, Hutton concluded that geological time must be very long indeed to account for the large changes apparent in the rock record. But this principle needs to be taken with a grain of salt: there are some processes that have occurred in the past that are no longer occurring (e.g., eruption of ultramafic lavas), as well as some processes that occur so irregularly that we have not yet witnessed such an event in historic time (e.g., impact of a large asteroid with Earth).

William Smith worked as a surveyor in the coal-mining and canal-building industries in south-western England in the late 1700s and early 1800s. While doing his work, he had many opportunities to observe Paleozoic and Mesozoic sedimentary rocks of the region, and he did so in a way that few had done before. Smith noticed the textural similarities and differences between rocks in different locations. More importantly, he discovered that fossils could be used to correlate rocks of the same age. Smith is credited with formulating the **principle of faunal succession**, the concept that specific types of organisms lived during different time intervals. He used the principle of faunal succession to great effect in his monumental project to create a geological map of England and Wales, published in 1815.

Inset into Smith's great geological map is a small diagram showing a schematic geological cross-section extending from the Thames estuary of eastern England to the west coast of Wales. Smith showed the sequence of rocks, from the Paleozoic rocks of Wales and western England, through the Mesozoic rocks of central England, to the Cenozoic rocks of the area around London (Figure 19.2).

Figure 19.2 William Smith's "Sketch of the succession of strata and their relative altitudes," an inset on his geological map of England and Wales (with era names added). *Source: Steven Earle (2015) CC BY 4.0 view source, modified after William Smith (1815) Public Domain*

Smith did not put any dates on these rocks, because he didn't know them. But he was aware of the **principle of superposition**, the idea developed much earlier by the Danish theologian and scientist Nicholas Steno, that young sedimentary rocks form on top of older ones. Therefore, Smith knew that this diagram represented a stratigraphic column. And since almost every period of the Phanerozoic is represented along this section through Wales and England, it is also a primitive geological time scale.

Smith's work set the stage for the naming and ordering of the geological time periods, which was initiated around 1820, first by British geologists, and later by other European geologists. Many of the periods are named for places where rocks of that age are found in Europe, such as Cambrian for Cambria in Wales, Devonian for Devon in England, Jurassic for the Jura Mountains in France and Switzerland, and Permian for the Perm region of Russia. Some are named for the type of rock that is common during that age, such as Carboniferous for the coal-bearing rocks of England, and Cretaceous for the chalks of England and France.

The early time scales were only relative because 19th century geologists did not know the absolute ages of rocks. This information was not available until the development of isotopic dating techniques early in the 20th century.

The geological timescale is currently maintained by the International Commission on Stratigraphy (ICS), which is part of the International Union of Geological Sciences. The timescale is updated continuously as we learn more about the timing and nature of past geological events. View the ICS timescale at http://www.stratigraphy.org/index.php/ics-chart-timescale

Geological time has been divided into four eons: Hadean, Archean, Proterozoic, and Phanerozoic (Figure 19.3). The first three of these eons represent almost 90% of Earth's history. Rocks from the Phanerozoic (meaning "visible life") are the most commonly exposed rocks on Earth, and they contain evidence of life forms with which we are familiar.

Figure 19.3 The eons of Earth's history. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0*

The Phanerozoic — the past 541 Ma of Earth's history — is divided into three eras: the Paleozoic ("early life"), the Mesozoic ("middle life"), and the Cenozoic ("new life"), and each era is divided into periods (Figure 19.4). Most of the organisms with which we share Earth evolved into familiar forms at various times during the Phanerozoic.

Figure 19.4 The eras (middle row) and periods (bottom row) of the Phanerozoic. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0*

The Cenozoic, representing the past 66 Ma, is divided into three periods, the Paleogene, Neogene, and Quaternary; and seven epochs (Figure 19.5). Non-avian dinosaurs became extinct at the start of the Cenozoic, after which birds and mammals radiated to fill the available habitats. Earth was very warm during the early Eocene, and has cooled steadily ever since. Glaciers first appeared on Antarctica in the Oligocene and then on Greenland in the Miocene. By the Pleistocene, glaciers covered much of North America and Europe. The most recent of the Pleistocene glaciations ended ~11,700 years ago. The current epoch is known as the Holocene. Epochs are further divided into ages.

Figure 19.5 The periods and epochs of the Cenozoic Era. *Source: Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0*

Most of the boundaries between the periods and epochs of the geological timescale have been fixed on the basis of significant changes in the fossil record. For example, the boundary between the Cretaceous and the Paleogene coincides exactly with the extinction of the dinosaurs. This is not a coincidence. Many other types of organisms went extinct at this time, and the boundary between the two periods marks the division between sedimentary rocks containing Cretaceous organisms below, and those containing Paleogene organisms above.

19.2 Relative Dating Methods

Relative Dating Principles

The simplest and most intuitive way of dating geological features is to look at the relationships between them. There are a few simple rules for doing this. But caution must be taken, as there may be situations in which the rules are not valid, so local factors must be understood before an interpretation can be made. These situations are generally rare, but they should not be forgotten when unraveling the geological history of an area.

The **principle of superposition** states that sedimentary layers are deposited in sequence, and the layers at the bottom are older than those at the top. This situation may not be true, though, if the sequence of rocks has been flipped completely over by tectonic processes, or disrupted by faulting.

The **principle of original horizontality** indicates that sediments are originally deposited as horizontal to nearly horizontal sheets. At a broad scale this is true, but at a smaller scale it may not be. For example, cross-bedding forms at an appreciable angle, where sand is deposited upon the lee face of a ripple. The same holds true of delta foreset beds (Figure 19.6).

Figure 19.6 A cross-section through a river delta forming in a lake. The delta foresets are labeled "Delta deposits" in this figure, and you can quickly see that the front face of the foresets are definitely not deposited horizontally. *Source: AntanO (2017) CC BY 4.0*

The **principle of lateral continuity** states that sediments are deposited such that they extend laterally for some distance before thinning and pinching out at the edge of the depositional basin. But sediments can also terminate against faults or erosional features (see unconformities below), so may be cut off by local factors.

The **principle of inclusions** states that any rock fragments that are included in a rock must be older than the rock in which they are included. For example, a **xenolith** in an igneous rock, or a clast in sedimentary rock must be older than the rock that includes it (Figure 19.7). A possible situation that would violate this principle is the following: an igneous dyke may intrude through a sequence of rocks, hence is younger than these rocks (see the principle of cross-cutting relationships below). Later deformation may cause the dyke to be pulled apart into small pieces, surrounded by the host rocks. This situation can make the pieces of the dyke appear to be xenoliths, but they are *younger* than the surrounding rock in this case.

Figure 19.7 Applications of the principle of inclusion. Left- A xenolith of diorite incorporated into a basalt lava flow, Mauna Kea volcano, Hawai'i. The lava flow took place some time after the diorite crystallized (hammer head for scale). Right- Rip-up clasts of shale embedded in Gabriola Formation sandstone, Gabriola Island, BC. The pieces of shale were eroded as the sand was deposited, so the shale is older than the sandstone. *Source: Karla Panchuk (2018) CC BY 4.0. Photographs by Steven Earle (2015) CC BY 4.0 view sources*

The **principle of cross-cutting relationships** states that any geological feature that cuts across or disrupts another feature must be younger than the feature that is disrupted. An example of this is given in Figure 19.8, which shows three different sedimentary layers. The lower sandstone layer is disrupted by two faults, so we can infer that the faults are younger than this layer. But the faults do not appear to continue into the coal seam, and they certainly do not continue into the upper sandstone. So we can infer that coal seam is younger than the faults (because the coal seam cuts across them). The upper sandstone is youngest of all, because it lies on top of the coal seam. An example that violates this principle can be seen with a type of fault called a growth fault. A growth fault is a fault that continues to move as sediments are continuously delivered to the hangingwall block. In this case, the lower portion of the fault that cuts the lower sediments may have originally formed *before* the uppermost sediments were deposited, despite the fault cutting through all of the sediments, and appearing to be entirely younger than all of the sediments.

Figure 19.8 Superposition and cross-cutting relationships in Cretaceous Nanaimo Group rocks in Nanaimo BC. The coal seam is about 50 cm thick. *Source: Steven Earle (2015) CC BY 4.0*

The **principle of baked contacts** states that the heat of an intrusion will bake (metamorphose) the rocks in close proximity to the intrusion. Hence the presence of a baked contact indicates the intrusion is younger than the rocks around it. If an intrusive igneous rock is exposed via erosion, then later buried by sediments, the surrounding rocks will *not* be baked, as the intrusion was already cold at the time of sediment deposition. But baked contacts may be difficult to discern, or may be minimally developed to absent when the intrusive rocks are low in volume or felsic (relatively cool) in composition.

The **principle of chilled margins** states that the portion of an intrusion that has cooled and crystallized next to cold surrounding rock will form smaller crystals than the portion of the intrusion that cooled more slowly deeper in the instrusion, which will form larger crystals. Smaller crystals generally appear darker in colour than larger crystals, so a chilled margin appears as a darkening of the intrusive rock towards the surrounding rock. This principle can be used to distinguish between an igneous sill, which will have a chilled margin at top and bottom, and a subaerial lava flow, which will have a chilled margin only at the bottom.

Exercise: Cross-Cutting Relationships

The outcrop in Figure 19.9 has three main rock types:

1. Buff/pink felsic intrusive igneous rock present as somewhat irregular masses trending from lower right to upper left
2. Dark grey metamorphosed basalt
3. A 50 cm wide light-grey felsic intrusive igneous dyke extending from the lower left to the upper right – offset in several places

Use the principle of cross-cutting relationships to determine the relative ages of these three rock types.

Figure 19.9 Outcrop from Horseshoe Bay, BC. *Source: Steven Earle CC BY 4.0*

Note: The near-vertical stripes are blasting drill holes. The image is about 7 m across.

Unconformities

An **unconformity** represents an interruption in the process of deposition of sediments. Recognizing unconformities is important for understanding time relationships in sedimentary sequences. An unconformity is visible in the Grand Canyon (Figure 19.10, white dashed line) above Proterozoic rocks that were tilted and then eroded to a flat surface prior to deposition of the younger Paleozoic rocks. The difference in time between the youngest of the Proterozoic rocks and the oldest of the Paleozoic rocks is

nearly 300 million years. Tilting and erosion of the older rocks took place during this time, and if there were any deposition occurring in this area during this time, erosion removed those sediments.

19.10 Angular unconformity in the Grand Canyon in Arizona, with a sketch of rock orientations. The tilted rocks at the bottom are part of the Proterozoic Grand Canyon Group (aged 825 to 1,250 Ma). The flat-lying rocks at the top are Paleozoic (540 to 250 Ma). The boundary between the two (dashed white line) represents a time gap of nearly 300 million years. *Source: Karla Panchuk (2018) CC BY 4.0. Photograph by Steven Earle (2015) CC BY 4.0*

There are four types of unconformities, reflecting different arrangements and types of rocks above and below the surface of non-deposition or erosion (Table 19.1, Figure 19.11).

Table 19.1 Characteristics of the Four Types of Unconformities

Unconformity Type	Description
Nonconformity	A boundary between non-sedimentary rocks below and sedimentary rocks above (Figure 19.11a)
Angular Unconformity	A boundary between two sequences of sedimentary rocks where the underlying units have been tilted (or folded) and eroded prior to the deposition of the younger units (Figures 19.10, 19.11b)
Disconformity	A boundary between two sequences of sedimentary rocks where the underlying units have been eroded (but not tilted) prior to the deposition of the younger units (Figures 19.8, 19.11c)
Paraconformity	A time gap in a sequence of sedimentary rocks due to non-deposition. The time gap does not show up as an angular unconformity or a disconformity (Figure 19.11d).

Figure 19.11 The four types of unconformities: (a) a nonconformity between non-sedimentary rock and sedimentary rock, (b) an angular unconformity, (c) a disconformity between layers of sedimentary rock, where the older rock has been eroded but not tilted, and (d) a paraconformity where there is a long period (millions of years) of non-deposition between two parallel layers. *Source: Steven Earle (2015) CC BY 4.0*

Exercise: Relative Dating with Unconformities

1. The surfaces G and H in Figure 19.12 are unconformities. What kind?
2. If erosion at the surface stopped and sediments were deposited once again, what kind of unconformity would exist between the layer I and younger rocks?
3. Provide a list of the events that affected the rocks in Figure 19.12 in order from the oldest event to the most recent event. Note that C and D are faults. The sedimentary rocks labelled A are folded, but the other sedimentary rocks are horizontal.

Figure 19.12 Block diagram showing sedimentary and igneous rocks affected by faults, folds, and erosion. *Source: Karla Panchuk (2018) CC BY-SA 4.0, modified after Woudloper (2009) CC BY-SA 1.0*

19.3 Dating Rocks Using Fossils

Geologists obtain a wide range of information from fossils. They help us to understand evolution, and life in general; they provide critical information for understanding depositional environments and changes in Earth's climate; and they can be used to date rocks.

Although the recognition of fossils goes back hundreds of years, the systematic cataloguing and assignment of relative ages to different organisms from the distant past—paleontology—only dates back to the earliest part of the 19th century. The oldest undisputed fossils are from rocks dated ~3.5 Ga, and although fossils this old are typically poorly preserved and are not useful for dating rocks, they can still provide important information about conditions at the time. The oldest well-understood fossils are from rocks dating back to ~600 Ma, and the sedimentary record from this time forward is rich in fossil remains that provide a detailed record of the history of life. However, as anyone who has gone hunting for fossils knows, this does not mean that all sedimentary rocks have visible fossils or that they are easy to find. Fossils alone cannot provide us with numerical ages of rocks, but over the past century geologists have acquired enough isotopic dates from rocks associated with fossiliferous rocks (such as igneous dykes cutting through sedimentary layers) to be able to put specific time limits on most fossils.

A selective history of life on Earth over the past 600 million years is provided in Figure 19.13. The major groups of organisms that we are familiar with appeared between the late Proterozoic and the Cambrian (~600 Ma to ~541 Ma). Plants, which originally evolved in the oceans as green algae, invaded land during the Ordovician (~450 Ma). Insects, which evolved from marine arthropods, invaded land during the Devonian (400 Ma), and amphibians (i.e., vertebrates) invaded land about 50 million years later. By the late Carboniferous, trees had evolved from earlier plants, and reptiles had evolved from amphibians. By the mid-Triassic, dinosaurs and mammals had evolved from reptiles and reptile ancestors, Birds evolved from dinosaurs during the Jurassic. Flowering plants evolved in the late Jurassic or early Cretaceous. The earliest primates evolved from other mammals in the early Paleogene, and the genus *Homo* evolved during the late Neogene (~2.8 Ma).

Figure 19.13 A selective summary of life on Earth during the late Proterozoic and the Phanerozoic. The top row shows geological eras, and the lower row shows geological periods. *Source: Steven Earle (2015) CC BY 4.0*

If we understand the sequence of evolution on Earth, we can apply this knowledge to determining the relative ages of rocks. This is William Smith's **principle of faunal succession**, although in spite of the name, it can apply to fossils of plants and simple organisms as well as to fauna (animals).

The Phanerozoic Eon has witnessed five major extinctions (stars in Figure 19.13). The most significant of these was at the end of the Permian, which saw the extinction of over 80% of all species, and over 90% of all marine species. Most well-known types of organisms that survived were still severely impacted by this event. The second most significant extinction occurred at the Cretaceous-Paleogene boundary (K-Pg, also known the Cretaceous-Tertiary or K-T extinction). At that time, ~75% of marine species disappeared, as well as dinosaurs (but not birds) and pterosaurs. Other species were badly reduced but survived, and then flourished in the Paleogene. The K-Pg extinction may have been caused by the impact of a large asteroid (10 km to 15 km in diameter) and/or volcanic eruptions associated with the formation of the Deccan Traps, but it is generally agreed that the other four Phanerozoic mass extinctions had other causes, although their exact nature is not clearly understood.

It is not a coincidence that the major extinctions all coincide with boundaries of geological periods and/or eras. Paleontologists have placed most of the divisions of the geological time scale at points in the fossil record where there are major changes in the type of fossils observed.

If we can identify a fossil, and we know when the organism lived, we can assign a range of time to the formation of the sediments in which the organism was preserved when it died. This range might be several millions of years, because some organisms survived for a very long time. If the rock we are studying has several types of fossils in it, and we can assign time ranges to all of these fossils, we may be able to narrow the time range for the age of the rock considerably (Figure 19.14).

Figure 19.14 Application of bracketing to constrain the age of a rock based on the presence of several fossils. The yellow bar represents the time range during which each of the four species (A – D) existed on Earth. Although each species lived for several millions of years, we can narrow down the age of the rock to a span of just 1.3 Ma during which all four species coexisted. *Source: Steven Earle (2015) CC BY 4.0*

Some organisms survived for a very long time, and are not particularly useful for dating rocks. Sharks, for example, have existed for over 400 million years, and the great white shark has survived for 16 million years so far. Organisms that lived for relatively short time periods are particularly useful for dating rocks, especially if they were distributed over a wide geographic area and hence can be used to compare rocks from different regions. These are known as **index fossils**. There is no specific limit on how short the time span has to be for a fossil to qualify as an index fossil. Some such organisms lived for millions of years, and others for much less than a million years.

Some well-studied groups of organisms qualify as **biozone** fossils because, although the genera and families lived over a long time, each species lived for a relatively short time and can be easily distinguished from others on the basis of specific features. For example, ammonites have a distinctive feature known as the **suture** line, where the internal shell layers that separate the individual chambers (**septae**) meet the outer shell wall (Figure 19.15). These suture lines are sufficiently variable to identify species that can be used to estimate the relative or absolute ages of the rocks in which they are found.

Figure 19.15 The septum of an ammonite (white part, left), and the suture lines where the septae meet the outer shell (right). *Source: Steven Earle (2015) CC BY 4.0*

Foraminifera—small, calcium carbonate-shelled marine organisms that originated during the Triassic and are still alive today—are also useful biozone fossils. Numerous different foraminifera lived during the Cretaceous Period, some for over 10 million years, but others for less than 1 million years (Figure 19.16). If the foraminifera in a rock can be identified to the species level, the rock's age can be determined.

Figure 19.16 Time ranges for Cretaceous foraminifera (left), and modern foraminifera from the Ambergris area of Belize (right). *Source: Left- Steven Earle (2015) CC BY 4.0, from data in Scott (2014). Right- Steven Earle (2015) CC BY 4.0*

Exercise: Dating Rocks Using Index Fossils

Figure 19.17 shows the age ranges for some late Cretaceous inoceramid clams in the genus *Mytiloides*. Using the bracketing method described above, determine the possible age range of a rock in which all five of these organisms were found.

How would the age range change if *M. subhercynius* were not present in the rock?

Figure 19.17 Inoceramid ranges. *Source: Steven Earle (2015) CC BY 4.0, from data in Harries et al. (1996).*

Correlation

Geologists employ relative age dating techniques to **correlate** rocks between regions. Correlation seeks to relate the geological history between regions, by relating the rocks in one region to those in another.

There are different techniques of correlation. The easiest technique is to correlate by rock type, or lithology, called **lithostratigraphic correlation**. In this method, specific rock types are related between regions. If a sequence of rocks at one site consists of a sandstone unit overlain by a limestone unit, then a unit of shale, and the exact same sequence of rocks—sandstone, limestone, shale—occurs at a nearby site, lithostratigraphic correlation means assuming that the rocks at both sites are in the same rocks. If you could see all of the rock exposed between the two sites, the units would connect with one another. The problem with this type of correlation is that some rocks may only have formed locally, or may pinch out between the two sites, and therefore not be present at the site to which a correlation is being attempted.

Another technique, **biostratigraphic correlation**, involves correlation based on fossil content. This technique uses fossil assemblages (fossils of different organisms that occur together) to correlate rocks between regions. The best fossils to use are those that are widely spread, abundant, and lived for a relatively short period of time.

Yet another technique, **chronostratigraphic correlation**, is to correlate rocks that have the same age. This can be the most difficult way to correlate, because rocks are generally diachronous. That is, if we trace

a given rock unit across any appreciable lateral distance, the age of that rock actually changes. To give a familiar example, when you go to the beach, you know that the beach itself and the lake bottom in the shallow water is sandy. But if you swim out to deeper water and touch bottom, the bottom feels muddy. The difference in sediment type has to do with the energy of deposition, with the waves at and near the beach keeping any fine sediments away, only depositing them in deeper quieter waters. If you think of this example in time, you realize that the sand at and near the beach is being deposited at the same time as the mud in deeper water. But if lake levels drop, the beach sands will slowly migrate outwards and cover some of the deeper water muds. If lake levels rise, the deeper water muds will slowly migrate landwards and cover some of the shallower water sands. This is an example of **Walther's Law**, which states that sedimentary rocks that we see one on top of each other in the rock record actually formed adjacent to one another at the time of deposition. In order to correlate rock units in time, we must target marker beds that formed instantaneously. An example of such would be an ash layer from a nearby volcano that erupted and blanketed an entire region in ash. But such marker beds are usually rare to absent, making such correlation extremely difficult.

19.4 Isotopic Dating Methods

Isotope Pairs

Originally, fossils only provided us with relative ages because, although early paleontologists understood biological succession, they did not know the absolute ages of the different organisms. It was only in the early part of the 20th century, when isotopic dating methods were first applied, that it became possible to discover the absolute ages of the rocks containing fossils. In most cases, we cannot use isotopic techniques to directly date fossils or the sedimentary rocks in which they are found, but we can constrain their ages by dating igneous rocks that cut across sedimentary rocks, or volcanic ash layers that lie within sedimentary layers.

Isotopic dating of rocks, or the minerals within them, is based upon the fact that we know the decay rates of certain unstable isotopes of elements, and that these decay rates have been constant throughout geological time. It is also based on the premise that when the atoms of an element decay within a mineral or a rock, they remain trapped in the mineral or rock, and do not escape.

One of the isotope pairs commonly used to date rocks is the decay of ^{40}K to ^{40}Ar (potassium-40 to argon-40). ^{40}K is a radioactive isotope of potassium that is present in very small amounts in all minerals that contain potassium. It has a **half-life** of 1.3 billion years, meaning that over a period of 1.3 Ga one-half of the ^{40}K atoms in a mineral or rock will decay to ^{40}Ar. Over the next 1.3 Ga one-half of the remaining atoms will decay, and so forth (Figure 19.18). ^{40}K is called the **parent isotope**, and ^{40}Ar the **daughter isotope**, as the parent transforms into the daughter during radioactive decay.

In order to use the K-Ar dating technique, we need to have an igneous or metamorphic rock that includes a potassium-bearing mineral. One good example is granite, which contains the mineral potassium feldspar (Figure 19.19). Potassium feldspar does not contain any argon when it forms. Over time, the ^{40}K in the feldspar decays to ^{40}Ar. The atoms of ^{40}Ar remain embedded within the crystal, unless the rock is subjected to high temperatures after it forms. The sample must be analyzed using a very sensitive mass-spectrometer, which can detect the differences between the masses of atoms, and can therefore distinguish between ^{40}K and the much more abundant ^{39}K. The minerals biotite and hornblende are also commonly used for K-Ar dating.

Figure 19.18 The decay of ^{40}K over time. Each half-life is 1.3 billion years, so after 3.9 billion years (three half-lives) 12.5% of the original ^{40}K will remain. The red-blue bars represent ^{40}K and the green-yellow bars represent ^{40}Ar. *Source: Steven Earle (2015) CC BY 4.0*

Figure 19.19 Crystals of potassium feldspar (pink) in a granitic rock are candidates for isotopic dating using the K-Ar method because they contained potassium and no argon when they formed. *Source: Steven Earle (2015) CC BY 4.0*

There are many isotope pairs that can be employed in dating igneous and metamorphic rocks (see Table 19.2), each with its strengths and weaknesses. In the above example, the daughter isotope ^{40}Ar is naturally a gas, and can escape the potassium feldspar quite easily if the feldspar is exposed to heating during metamorphism, or interaction with hydrothermal fluids. Hence we must closely examine the feldspar mineral to determine if there is any evidence of alteration. If some ^{40}Ar has been lost, but the sample is dated anyway, an age that is too young will be calculated.

Table 19.2 Commonly Used Isotope Systems for Dating Geological Materials

Isotope System	Half-life	Useful Range	Comments
Potassium-Argon	1.3 Ga	10 ka - 4.57 Ga	Widely applicable because most rocks contain potassium
Uranium-Lead	4.5 Ga	1 Ma - 4.57 Ga	The rock must contain uranium-bearing minerals (felsic igneous rocks)
Rubidium-Strontium	47 Ma	10 Ma - 4.57 Ga	Less precision than other methods for old rocks

Carbon-Nitrogen (radiocarbon dating)	5,730 a	100 a to 60 ka	Sample must contain wood, bone, or carbonate minerals; can be applied to young sediments

Each parent isotope has a certain half-life, which ranges from microseconds to billions of years, depending upon the isotope. In dating rocks, we need to select an isotope pair with a parent isotope that has a reasonable half-life. This means that the half-life must not be too short or too long. If the half-life is too short, then most of the parent isotope will have decayed to form the daughter isotope. If we cannot measure the amount of parent isotope very accurately, which will be impossible to do if there is only the tiniest amount of parent isotope left, the calculated age will have huge errors associated with it. The same applies if the half-life is too long. In this case, very little of the daughter isotope will have formed, and our inability to measure the small amount of daughter isotope accurately will again result in huge errors in the calculated age.

Another complicating factor is whether the mineral of interest incorporated any of the daughter isotope into its structure at the time of formation. When we select a mineral and an isotope pair to date that mineral, we make the assumption that all of the daughter isotope we find in the mineral was produced in the mineral by radioactive decay of the parent isotope. But if the mineral formed with some daughter isotope already present in its structure, then the age we calculate will be too old.

A more robust mineral to use to date certain types of igneous and metamorphic rocks is zircon. Zircon is a mineral that incorporates uranium into its structure at the time of formation. One of the isotopes of uranium decays to lead with a long half-life (see Table 19.2). Zircon is a mineral of choice for dating because it takes no lead into its structure when it forms, so any lead present is due entirely to the radioactive decay of the uranium parent. Another reason is because zircon is a very resistant mineral. It can handle exposure to hydrothermal fluids, and all but the highest grades of metamorphism, and not lose any of the parent or daughter isotopes. Hence the age that we calculate tends to be very accurate. One drawback is that zircon tends to form only in felsic igneous rocks. If we are trying to date a mafic igneous rock, we must choose a different mineral.

The Meaning of a Radiometric Date

When we employ isotopic methods on minerals, we are measuring an **age date**. Generally, an age date refers to the time since a mineral crystallized from molten rock (magma or lava). This is when the elements that make up the mineral get locked into the mineral's structure. But as we have already seen, elevated temperatures can cause elements to escape from a mineral, without the mineral melting. Hence when we date a mineral, we may be dating the time since the mineral crystallized from a melt, *or* the time since the mineral last experienced a period of heating above its **Curie point**, which is the temperature beyond which the mineral is able to lose (or gain) elements from its structure, without melting. So we have to know something about the rock before we forge ahead to measure an age. We may choose a mineral and isotope pair that are very resistant to metamorphism, so that we can "see through" the metamorphism, and determine the original age that the mineral crystallized from a melt. Or we may be interested in the age of the metamorphic event itself, so choose a mineral and isotope pair that is susceptible to resetting the isotopic clock during metamorphism (such as by losing all of the daughter isotope).

Absolute age dating is a powerful tool for unraveling the geological history of a region, but we must ultimately rely upon igneous rocks (that may have later metamorphosed) for the minerals that we are able to date (see the next section for issues with dating sedimentary rocks directly). Another issue with absolute age dating is that it is expensive, with a single analysis costing several hundreds of dollars. This means geologists never forget their relative age dating principles, and are always applying them in the field to determine the sequence of events that formed the rocks in a region.

Exercise: Combining Absolute Ages with Relative Dating

The age dates for three igneous rock layers are given. Use relative dating techniques to determine the age ranges for the sets of sedimentary units A, B, and C.

Figure 19.20 A sequence of igneous and sedimentary layers. Age dates are given for the igneous layers. *Source: Karla Panchuk (2018) CC BY-SA 4.0, modified after Jill Curie (2014) CC BY-SA 3.0*

Isotope Dating Techniques and Sedimentary Rocks

A clastic sedimentary rock (e.g., conglomerate, Figure 19.21) is made up of older rock and mineral fragments. These fragments were derived from weathering and erosion of pre-existing rocks. The process of forming a sedimentary rock from sediments generally occurs at low temperatures. The minerals are not heated beyond their Curie points, and therefore still preserve their original ages (either igneous crystallization age, or a metamorphic age).

In almost all cases, the fragments have come from a range of source rocks that all formed at different times. If we dated a number of individual grains in the sedimentary rock, we would likely get a range of different dates, all older than the age of the sedimentary rock. The most that such ages gleaned from a sedimentary rock can tell us is a maximum age of the sedimentary rock. It might be possible to date some chemical sedimentary rocks isotopically, but there are no useful isotopes that can be used on old chemical sedimentary rocks.

Figure 19.21 Conglomerate is a sedimentary rock consisting of large rounded clasts surrounded by finer-grained material. *Source: Steven Earle (2015) CC BY 4.0*

Radiocarbon Dating

Radiocarbon dating (using ^{14}C) can be applied to many geological materials, including sediment and sedimentary rocks, but only if the materials in question are younger than ~60 ka, and contain organic material. Beyond this time, so little ^{14}C is left that it cannot be measured accurately, and the resulting age dates are unreliable. Fragments of wood incorporated into young sediment are good candidates for carbon dating, and this technique has been used widely in studies involving late Pleistocene glaciers and glacial sediments. Figure 19.22 shows radiocarbon dates from wood fragments in glacial sediments have been used to estimate the time of the last glacial advance along the Strait of Georgia.

Figure 19.22 Radiocarbon dates on wood fragments in glacial sediments in the Strait of Georgia. *Source: Steven Earle (2015) CC BY 4.0, modified after Clague (1976).*

Exercise: Radiometric Dating with Potassium-Argon

Assume that a feldspar crystal from the granite shown in Figure 19.19 was analyzed for ^{40}K and ^{40}Ar. The proportion of ^{40}K remaining is 0.91. Using the decay curve shown on this graph, estimate the age of the rock. An example is provided (in blue) for a ^{40}K proportion of 0.95, which is equivalent to an age of approximately 96 Ma. This is determined by drawing a horizontal line from 0.95 to the decay curve line, and then a vertical line from there to the time axis.

Figure 19.23 Decay curve for potassium-argon dating. *Source: Steven Earle (2015) CC BY 4.0*

19.5 Other Dating Methods

There are numerous other techniques for dating geological materials, but we will examine just two of them here: dendrochronology—tree-ring dating—and dating based on the record of reversals of Earth's magnetic field.

Dendrochronology

Dendrochronology can be applied to dating very young geological materials based on reference records of tree-ring growth going back many millennia. The longest such records can take us back over 25 ka, to the height of the last glaciation. One of the advantages of dendrochronology is that, providing reliable reference records are available, the technique can be used to date events to the nearest year.

Dendrochronology has been used to date the last major subduction zone earthquake on the coast of B.C., Washington, and Oregon. When large earthquakes occur in this region, there is a tendency for some coastal areas to subside by one or two metres. Seawater then rushes in, flooding coastal flats and killing trees and other vegetation within a few months. There are at least four locations along the coast of Washington that have such dead trees, and probably many more in other areas. Wood samples from these trees have been studied and the ring patterns have been compared with patterns from old living trees in the region (Figure 19.24).

At all of the locations studied, the trees were found to have died either in the year 1699, or very shortly thereafter (Figure 19.25). On the basis of these results, it was concluded that a major earthquake took place in this region sometime between the end of growing season in 1699 and the beginning of the growing season in 1700. Evidence from a major tsunami that struck Japan on January 27, 1700, narrowed the timing of the earthquake to sometime in the evening of January 26, 1700. (For more information, see https://web.viu.ca/earle/1700-quake/.)

Top- **Figure 19.24** Example of tree-ring dating of dead trees. *Source: Steven Earle (2015) CC BY 4.0*

Right- **Figure 19.25** Sites in Washington where dead trees are present in coastal flats. The outermost wood of eight trees was dated using dendrochronology, and of these, seven died during the year 1699, suggesting that the land was inundated by water at this time. *Source: Steven Earle (2015) CC BY 4.0, from data in Yamaguchi et al. (1997).*

Magnetic Chronology

Changes in Earth's magnetic field can also be used to date events in geologic history. The magnetic field causes compass needles point toward the north magnetic pole, but this hasn't always been the case. At various times in the past, Earth's magnetic field has reversed its polarity, and during such times a compass needle would have pointed to the south magnetic pole. By studying magnetism in volcanic rocks that have been dated isotopically, geologists have been able to establish the chronology of magnetic field reversals going back for ~250 Ma. About 5 Ma of this record is shown in Figure 19.26, where the black bands

represent periods of normal magnetism ("normal" meaning a polarity identical to the current magnetic field) and the white bands represent periods of reversed magnetic polarity. These periods of consistent magnetic polarity are given names to make them easier to reference. The current period of normal magnetic polarity, known as Brunhes, has lasted for the past 780,000 years. Prior to that there was a short reversed period and then a short normal period, the latter of which is known as Jaramillo.

Figure 19.26 The last 5 Ma of magnetic field reversals. *Source: Steven Earle (2015) CC BY 4.0 view source, modified after U.S. Geological Survey (2007) Public Domain*

Oceanic crust becomes magnetized by the magnetic field that exists as the crust forms from magma at mid-ocean ridges. As it cools, the magnetic fields of tiny crystals of magnetite that form within the magma become aligned with the existing magnetic field, and remain in this orientation, even if Earth's magnetic field later changes polarity (Figure 19.27). Oceanic crust that is forming today is being magnetized in a "normal" sense, but crust that formed 780,000 to 900,000 years ago, in the interval between the Brunhes and Jaramillo normal periods, was magnetized in the "reversed" sense.

Figure 19.27 Formation of magnetized oceanic crust at a spreading ridge. Coloured bars represent periods of normal magnetic polarity. Capital letters denote the Brunhes, Jaramillio, Olduvai, and Gauss normal magnetic periods (see Figure 19.26). *Source: Steven Earle (2015) CC BY 4.0*

Magnetic chronology can be used as a dating technique because we can measure the magnetic field of rocks using a magnetometer, or of entire regions by towing a magnetometer behind a ship or an airplane. For example, the Juan de Fuca Plate, which lies off of the west coast of BC, Washington, and Oregon, is being and has been formed along the Juan de Fuca spreading ridge (Figure 19.28). The parts of the plate that are still close to the ridge exhibit normal magnetic polarity, while parts that are further away (and formed much earlier) have either normal or reversed magnetic polarity, depending upon when the rock formed. By carefully matching the sea-floor magnetic stripes with the known magnetic chronology, we can determine the age at any point on the plate. We can see that the oldest part of the Juan de Fuca Plate that has not yet subducted (off of the coast of Oregon) is just over 8 million years old, while the part that is subducting beneath Vancouver Island is between 0 and ~6 million years old.

Figure 19.28 The pattern of magnetism within the area of the Juan de Fuca Plate, off the west coast of North America. Coloured bands represent parts of the sea floor with normal magnetic polarity, and the magnetic time scale is shown using these same colours. *Source: Steven Earle (2015) CC BY 4.0*

Exercise: Magnetic Dating

The fact that magnetic intervals can only be either *normal* or *reversed* places significant limits on the applicability of magnetic dating. If we find a rock with normal magnetism, we can't know which normal magnetic interval it represents, unless we have some other information.

Using Figure 19.26 for reference, determine the age of a rock with normal magnetism that has been found to be between 1.5 and 2.0 Ma based on fossil evidence in nearby sedimentary rocks.

How old is a rock that is limited to 2.6 to 3.2 Ma by fossils, and which has reversed magnetic polarity?

19.6 Understanding Geological Time

It is one thing to know the facts about geological time — how long it is, how we measure it, how we divide it into smaller time intervals, and what we call the various periods and epochs — but it is quite another to really *understand* geological time. The problem is that our lives are short and our memories are even shorter. Our experiences span only a few decades, so we really don't have a way of knowing what 11,700 years means. What's more, it is hard for us to understand how 11,700 years differs from 65.5 Ma, or even from 1.8 Ga. It is not that we cannot comprehend what the numbers mean, it is that we cannot really appreciate how much time is involved.

You may wonder why it is so important to understand geological time. There are some very good reasons. One is so that we can fully understand how geological processes that seem impossibly slow can produce anything of consequence. Consider driving from one major city to another, where a journey of several hours might occur at speeds of ~100 km/h. Continents move toward each other at rates of a fraction of a millimetre per day, a speed something on the order of 0.00000001 km/h, and yet, at this impossibly slow rate (try walking at this speed!), they can move thousands of kilometres through geological time. Sediments typically accumulate at even slower rates — less than a millimetre per year — but still they are thick enough to be thrust up to form huge mountains or carved into breathtaking canyons.

Another reason is to understand issues like extinction of endangered species, and human influence on climate. People who do not understand geological time are quick to say that the climate has changed in the past, and that what is happening now is no different. And climate certainly has changed in the past: from the Eocene (50 Ma) to the present day, Earth's climate cooled by ~12°C on average. This is a huge change that ranks as one of the most important climate changes of Earth's past, and yet the rate of change over this time was only 0.000024 °C/century. Recent warming has occurred at a rate of ~1.1°C over the past 100 years (NASA GISS), 45,800 times faster than the rate of climate change since the Eocene.

One way to wrap your mind around geological time is to put it into the perspective of single year. At this rate, each hour of the year is equivalent to approximately 500,000 years, and each day is equivalent to 12.5 million years. If all of geological time is compressed down into a single year, Earth formed on January 1, and the first life forms evolved in late March (~3,500 Ma). The first multicellular life forms appeared on November 13 (~600 Ma), plants appeared on land on November 24, and amphibians on December 3. Reptiles evolved from amphibians during the first week of December, and dinosaurs and early mammals evolved by December 13. Non-avian dinosaurs, which survived for 160 million years, went extinct on Boxing Day (December 26). The Pleistocene glaciation began at ~6:30 p.m. on New Year's Eve, and the last glacial ice melted from southern Canada by 11:59 p.m.

It is worth repeating: on this time scale, the earliest ancestors of the animals and plants with which we are familiar did not appear on Earth until mid-November, the dinosaurs disappeared after Christmas, and most of Canada was periodically locked in ice from 6:30 to 11:59 p.m. on New Year's Eve. As for people, the first to inhabit Canada arrived about one minute before midnight.

Exercise: What Happened on Your Birthday?

Using the "all of geological time compressed to one year" concept, determine the geological date that is equivalent to your birthday. First, go to http://mistupid.com/calendar/dayofyear.htm to find out which day of the year your birth date is. Divide that number by 365, and multiply that number by 4,570 to determine the time (in millions since the *beginning* of geological time). Subtract that number from 4,570 to determine the date back from the present.

Example: April Fool's Day (April 1) is day 91 of the year: 91/365 = 0.2493. 0.2493 x 4,570 = 1,139 million years from the start of time, and 4,570 - 1,193 = 3,377 Ma is the geological date.

Finally, go to the Foundation for Global Community's Walk through Time website at http://www.globalcommunity.org/wtt/walk_menu/ to find out what was happening on your day. The nearest date to 3,377 Ma is 3,400 Ma.

Chapter 19 Summary

The topics covered in this chapter can be summarized as follows:

19.1 The Geological Timescale

The work of William Smith was critical to the establishment of the first geological timescale early in the 19th century, but it wasn't until the 20th century that geologists were able to assign reliable dates to the various time periods. The geological timescale is now maintained by the International Commission on Stratigraphy. Geological time is divided into eons, eras, periods, and epochs.

19.2 Relative Dating Methods

We can determine the relative ages of different rocks by observing and interpreting relationships among them, such as superposition, cross-cutting, and inclusions. Gaps in the geological record are represented by various types of unconformities.

19.3 Dating Rocks Using Fossils

Fossils are useful for dating rocks back to ~600 Ma. If we know the age range of a fossil, we can date the rock in which it is found, but some organisms lived for many millions of years. Index fossils represent shorter geological time spans, and if a rock has several different fossils with known age ranges, we can narrow the time during which the rock formed.

19.4 Isotopic Dating Methods

Radioactive isotopes decay at constant known rates, and can be used to date igneous and metamorphic rocks. Some commonly used isotope systems are potassium-argon, rubidium-strontium, uranium-lead, and carbon-nitrogen. Radiocarbon dating can be applied to sediments and sedimentary rocks, but only if they are younger than 60 ka, and contain organic material, or minerals of calcium carbonate.

19.5 Other Dating Methods

There are many other methods for dating geological materials. Two that are widely used are dendrochronology and magnetic chronology. Dendrochronology, based on studies of tree rings, is widely applied to dating glacial events. Magnetic chronology is based on the known record of Earth's magnetic field reversals.

19.6 Understanding Geological Time

While understanding geological time is relatively easy, actually comprehending the significance of the vast amounts of geological time is a great challenge. To be able to solve important geological problems and certain societal challenges, we need to really appreciate the vastness of geological time.

Review Questions

1. A granitic rock contains inclusions (xenoliths) of basalt. What can you say about the relative ages of the granite and the basalt?
2. Explain the differences between a) disconformity and paraconformity; and b) nonconformity and angular unconformity
3. What are the features of a useful index fossil?
4. Figure 19.29 shows a geological cross-section. The granitic rock F at the bottom is the rock that you estimated the age of in the exercise in 19.4, *Radiometric Dating with Potassium-Argon*. A piece of wood from layer D has been sent for radiocarbon dating and the result was 0.55 ^{14}C remaining. How old is layer D?
5. Based on your answer to question 4, what can you say about the age of layer C in the figure above?
6. What type of unconformity exists between layer C and rock F?
7. What type of unconformity exists between layer C and rock B?

Figure 19.29 Geological cross-section (left) and decay curve for ^{14}C ages. *Source: Left- Karla Panchuk (2018) CC BY 4.0, modified after Steven Earle (2015) CC BY 4.0. Right- Steven Earle (2015) CC BY 4.0.*

8. We cannot use magnetic chronology to date anything older than ~780,000 years. Why?
9. How did William Smith apply the principle of faunal succession to determine the relative ages of the sedimentary rocks of England and Wales?
10. Access a copy of the geological time scale at http://www.stratigraphy.org/index.php/ics-chart-timescale. What are the names of the last age of the Cretaceous and the first age of the Paleogene? Print out the time scale and stick it on the wall above your desk!

Answers at the end of the chapter

References

Clague, J. (1976). Quadra Sand and its relation to late Wisconsin glaciation of southeast British Columbia. *Canadian Journal of Earth Sciences*, *13*, 803-815.

Harries, P.J., Kauffman, E.G., Crampton, J.S. (Redacteurs), Bengtson, P., Cech, S., Crame, J.A., Dhondt, A.V., Ernst, G., Hilbrecht, H., Lopez, Mortimore, G.R., Tröger, K.-A., Walaszcyk, I., & Wood, C.J. (1996). Mitteilungen aus dem Geologisch - Paläontologischen Museum der Universität Hamburg, *77*, 641-671. Retrieved from http://www.fuhrmann-hilbrecht.de/Heinz/geology/InoIntro/InoIntro.html

NASA Goddard Institute for Space Studies (n.d.). *GLOBAL Station Temperature Index in 0.01 degrees Celsius base period: 1951-1980* [data file]. Retrieved from http://data.giss.nasa.gov/gistemp/tabledata_v3/GLB.Ts.txt

Scott, R. (2014). A Cretaceous chronostratigraphic database: construction and applications, *Carnets de Géologie, 14*(2), 15-37. Retrieved from http://paleopolis.rediris.es/cg/1402/

Smith, W. (1815). A delineation of the strata of England and Wales with part of Scotland [map].

Yamaguchi, D.K., Atwater, B.F., Bunker, D.E., Benson, B.E., & Reid, M. S. (1997). Tree-ring dating the 1700 Cascadia earthquake. *Nature*, *389*, 922 - 923.

Answers to Review Questions

1. Xenoliths of basalt within a granite must be older than the granite according to the principle of inclusions.

2. (a) At both disconformities and paraconformities the beds above and below are parallel, but at a disconformity there is clear evidence of an erosion surface (the lower layers have been eroded). (b) A nonconformity is a boundary between sedimentary rocks above and non-sedimentary rocks below while an angular unconformity is a boundary between sedimentary rocks above and tilted and eroded and sedimentary layers below.
3. A useful index fossil must have survived for a relatively short period (e.g., around a million years), and also should have a wide distribution so that it can be used to correlate rocks from different regions.
4. The granitic rock F has been dated to 175 Ma. The wood in layer D is approximately 5,000 years old, so we can assume that layer D is no older than that, although it could be as much as a few hundred years younger if the wood was already old when it got incorporated into the rock.
5. Layer C must be between 5,000 y and 275 Ma.
6. The unconformity between layer C and rock F is a nonconformity.
7. The granite (F) was eroded prior to deposition of C, so it's likely that layer B was also eroded at the same time. If so, that makes the boundary between C and B a disconformity.
8. The last magnetic reversal was 780,000 years ago, so all rock formed since that time is normally magnetized and it isn't possible to distinguish older rock from younger rock within that time period using magnetic data.
9. William Smith was familiar with the different diagnostic fossils of the rocks of England and Wales and was able to use them to identify rocks of different ages.
10. The last age of the Cretaceous is the Maastrichtian (71.2 to 66.0 Ma) and the first age of the Paleogene is the Danian (66.0 to 61.6 Ma).

First Edition Chapters 14 & 21, by Steven Earle.

Chapter 14 Groundwater

Physical Geology, 1st Edition, by Steven Earle. https://opentextbc.ca/geology/

Introduction

Learning Objectives

After reading this chapter, completing the exercises within it, and answering the questions at the end, you should be able to:

- Explain the concepts of porosity and permeability and the importance of these to groundwater storage and movement
- Describe the relative porosities and permeabilities of some common geological materials
- Define aquifers, aquitards, confining layers, and the differences between confined and unconfined aquifers
- Explain the concepts of hydraulic head, the water table, potentiometric surface, and hydraulic gradient, and apply the Darcy equation for estimating groundwater flow
- Describe the flow of groundwater from recharge areas to discharge areas
- Describe the nature of groundwater flow in karst systems
- Explain how wells are used to extract groundwater and the implications of over-pumping a well
- Describe how observation wells are used to monitor groundwater levels and the importance of protecting groundwater resources
- Distinguish between natural and anthropogenic contamination of groundwater
- Describe some of the ways that groundwater can become contaminated, and how contamination can be minimized

Figure 14.1 A spring flowing from a limestone cave on Quadra Island, B.C. [SE]

As we saw in Chapter 13, fresh water makes up only 3% of the water on Earth. Approximately two-thirds of that is glacial ice and most of the rest is groundwater. We can't live without water,

and it's easy to see that groundwater represents a critically important component of our water supply. Groundwater is not as easily accessed as surface water, but it is also not as easily contaminated as surface water. If more than 7 billion of us want to continue living comfortably here on Earth, we have to take great care of our groundwater and learn how to use it sustainably.

14.1 Groundwater and Aquifers

Groundwater is stored in the open spaces within rocks and within unconsolidated sediments. Rocks and sediments near the surface are under less pressure than those at significant depth and therefore tend to have more open space. For this reason, and because it's expensive to drill deep wells, most of the groundwater that is accessed by individual users is within the first 100 m of the surface. Some municipal, agricultural, and industrial groundwater users get their water from greater depth, but deeper groundwater tends to be of lower quality than shallow groundwater, so there is a limit as to how deep we can go.

Porosity is the percentage of open space within an unconsolidated sediment or a rock. Primary porosity is represented by the spaces between grains in a sediment or sedimentary rock. Secondary porosity is porosity that has developed after the rock has formed. It can include fracture porosity — space within fractures in any kind of rock. Some volcanic rock has a special type of porosity related to vesicles, and some limestone has extra porosity related to cavities within fossils.

Porosity is expressed as a percentage calculated from the volume of open space in a rock compared with the total volume of rock. The typical ranges in porosity of a number of different geological materials are shown in Figure 14.2. Unconsolidated sediments tend to have higher porosity than consolidated ones because they have no cement, and most have not been strongly compressed. Finer-grained materials (e.g., silt and clay) tend to have greater porosity — some as high as 70% — than coarser materials (e.g., gravel). Primary porosity tends to be higher in well-sorted sediments compared to poorly sorted sediments, where there is a range of smaller particles to fill the spaces made by the larger particles. Glacial till, which has a wide range of grain sizes and is typically formed under compression beneath glacial ice, has relatively low porosity.

Consolidation and cementation during the process of lithification of unconsolidated sediments into sedimentary rocks reduces primary porosity. Sedimentary rocks generally have porosities in the range of 10% to 30%, some of which may be secondary (fracture) porosity. The grain size, sorting, compaction, and degree of cementation of the rocks all influence primary porosity. For example, poorly sorted and well-cemented sandstone and well-compressed mudstone can have very low porosity. Igneous or metamorphic rocks have the lowest primary porosity because they commonly form at depth and have interlocking crystals. Most of their porosity comes in the form of secondary porosity in fractures. Of the consolidated rocks, well-fractured volcanic rocks and limestone that has cavernous openings produced by dissolution have the highest potential porosity, while intrusive igneous and metamorphic rocks, which formed under great pressure, have the lowest.

Figure 14.2 Variations in porosity of unconsolidated materials (in red) and rocks (in blue) [SE]

Porosity is a measure of how much water can be stored in geological materials. Almost all rocks contain some porosity and therefore contain groundwater. Groundwater is found under your feet and everywhere on the planet. Considering that sedimentary rocks and unconsolidated sediments cover about 75% of the continental crust with an average thickness of a few hundred metres, and that they are likely to have around 20% porosity on average, it is easy to see that a huge volume of water can be stored in the ground.

Porosity is a description of how much space there could be to hold water under the ground, and **permeability** describes how those pores are shaped and interconnected. This determines how easy it is for water to flow from one pore to the next. Larger pores mean there is less friction between flowing water and the sides of the pores. Smaller pores mean more friction along pore walls, but also more twists and turns for the water to have to flow-through. A permeable material has a greater number of larger, well-connected pores spaces, whereas an impermeable material has fewer, smaller pores that are poorly connected. Permeability is the most important variable in groundwater. Permeability describes how easily water can flow through the rock or unconsolidated sediment and how easy it will be to extract the water for our purposes. The characteristic of permeability of a geological material is quantified by geoscientists and engineers using a number of different units, but the most common is the **hydraulic conductivity**. The symbol used for hydraulic conductivity is *K*. Although hydraulic conductivity can be expressed in a range of different units, in this book, we will always use m/s.

The materials in Figure 14.3 show that there is a wide range of permeability in geological materials from 10-12 m/s (0.000000000001 m/s) to around 1 m/s. Unconsolidated materials are generally more permeable than the corresponding rocks (compare sand with sandstone, for

example), and the coarser materials are much more permeable than the finer ones. The least permeable rocks are unfractured intrusive igneous and metamorphic rocks, followed by unfractured mudstone, sandstone, and limestone. The permeability of sandstone can vary widely depending on the degree of sorting and the amount of cement that is present. Fractured igneous and metamorphic rocks, and especially fractured volcanic rocks, can be highly permeable, as can limestone that has been dissolved along fractures and bedding planes to create solutional openings.

Figure 14.3 Variations in hydraulic conductivity (in metres/second) of unconsolidated materials (in red) and of rocks (in blue) [SE]

Why is clay porous but not permeable?

Both sand and clay deposits (and sandstone and mudstone) are quite porous (30% to 50% for sand and 40% to 70% for silt and clay), but while sand can be quite permeable, clay and mudstone are not.

The surface of most silicate mineral grains has a slight negative charge due to imperfections in the mineral structure. Water (H2O) is a polar molecule. This means that while it has no overall electrical charge, one side of the molecule has a slight positive charge (the side with the two hydrogens), compared to a slight negative charge on the other side. Water is strongly attracted to all mineral grains and water within that bound water layer (a few microns around each grain) is not able to move and flow along with the rest of the groundwater. In the lower diagrams shown here, the bound water is represented by dark blue lines around each grain and the water that can move is light blue. In the sand, there is still a lot of water that is able to move through the sediment, but in the clay/silt almost all of the water is held tightly to the grains and this reduces the permeability. [SE]

We have now seen that there is a wide range of porosity in geological materials and an even wider range of permeability. Groundwater exists everywhere there is porosity. However, whether that groundwater is able to flow in significant quantities depends on the permeability. An **aquifer** is defined as a body of rock or unconsolidated sediment that has sufficient permeability to allow water to flow through it. Unconsolidated materials like gravel, sand, and even silt make relatively good aquifers, as do rocks like sandstone. Other rocks can be good aquifers if they are well fractured. An **aquitard** is a body that does not allow transmission of a significant amount of water, such as a clay, a till, or a poorly fractured igneous or metamorphic rock. These are relative terms, not absolute, and are usually defined based on someone's desire to pump groundwater; what is an aquifer to someone who does not need a lot of water, may be an aquitard to someone else who does. An aquifer that is exposed at the ground surface is called an **unconfined aquifer**. An aquifer where there is a lower permeability material between the aquifer and the ground surface is known as a **confined aquifer**, and the aquitard separating ground surface and the aquifer is known as the **confining layer**.

Figure 14.4 shows a cross-section of a series of rocks and unconsolidated materials, some of which might serve as aquifers and others as aquitards or confining layers. The granite is much less permeable than the other materials, and so is an aquitard in this context. The yellow layer is very permeable and would make an ideal aquifer. The overlying grey layer is a confining layer.

The upper buff-coloured layer (*K* = 10-2 m/s) does not have a confining layer and is an unconfined aquifer. The yellow layer (*K* = 10-1 m/s) is "confined" by the confining layer (*K* = 10-4 m/s), and is a confined aquifer. The confined aquifer gets most of its water from the upper part of the hill where it is exposed at the surface, and relatively little by seepage through the fine silt layer.

Figure 14.4 A cross-section showing materials that might serve as aquifers and confining layers. The relative permeabilities are denoted by hydraulic conductivity (K = m/s). The pink rock is granite; the other layers are various sedimentary layers. [SE]

14.2 Groundwater Flow

If you go out into your garden or into a forest or a park and start digging, you will find that the soil is moist (unless you're in a desert), but it's not saturated with water. This means that some of the pore space in the soil is occupied by water, and some of the pore space is occupied by air (unless you're in a swamp). This is known as the **unsaturated zone**. If you could dig down far enough, you would get to the point where all of the pore spaces are 100% filled with water (saturated) and the bottom of your hole would fill up with water. The level of water in the hole represents the **water table**, which is the surface of the **saturated zone**. In most parts of British Columbia, the water table is several metres below the surface.

Water falling on the ground surface as precipitation (rain, snow, hail, fog, etc.) may flow off a hill slope directly to a stream in the form of **runoff**, or it may **infiltrate** the ground, where it is stored in the unsaturated zone. The water in the unsaturated zone may be used by plants (transpiration), evaporate from the soil (evaporation), or continue past the root zone and flow downward to the water table, where it **recharges** the groundwater.

A cross-section of a typical hillside with an unconfined aquifer is illustrated in Figure 14.5. In areas with topographic relief, the water table generally follows the land surface, but tends to come closer to surface in valleys, and intersects the surface where there are streams or lakes. The water table can be determined from the depth of water in a well that isn't being pumped, although, as described below, that only applies if the well is within an unconfined aquifer. In this case, most of the hillside forms the **recharge area**, where water from precipitation flows downward through the unsaturated zone to reach the water table. The area at the stream or lake to which the groundwater is flowing is a **discharge area**.

What makes water flow from the recharge areas to the discharge areas? Recall that water is flowing in pores where there is friction, which means it takes work to move the water. There is also some friction between water molecules themselves, which is determined by the viscosity. Water has a low viscosity, but friction is still a factor. All flowing fluids are always losing energy to friction with their surroundings. Water will flow from areas with high energy to those with low energy. Recharge areas are at higher elevations, where the water has high gravitational energy. It was energy from the sun that evaporated the water into the atmosphere and lifted it up to the recharge area. The water loses this gravitational energy as it flows from the recharge area to the discharge area.

In Figure 14.5, the water table is sloping; that slope represents the change in gravitational potential energy of the water at the water table. The water table is higher under the recharge area (90 m) and lower at the discharge area (82 m). Imagine how much work it would be to lift water 8 m high in the air. That is the energy that was lost to friction as the groundwater flowed from the top of the hill to the stream.

Figure 14.5 A depiction of the water table in cross-section, with the saturated zone below and the unsaturated zone above. The water table is denoted with a small upside-down triangle. [SE]

The situation gets a lot more complicated in the case of confined aquifers, but they are important sources of water so we need to understand how they work. As shown in Figure 14.6, there is always a water table, and that applies even if the geological materials at the surface have very low permeability. Where there is a confined aquifer — meaning one that is separated from the

surface by a confining layer — this aquifer will have its own "water table," which is actually called a **potentiometric surface**, as it is a measure of the total potential energy of the water. The red dashed line in Figure 14.6 is the potentiometric surface for the confined aquifer, and it describes the total energy that water is under within the confined aquifer. If we drill a well into the unconfined aquifer, the water will rise to the level of the water table (well A in Figure 14.6). But if we drill a well through both the unconfined aquifer and the confining layer and into the confined aquifer, the water will rise above the top of the confined aquifer to the level of its potentiometric surface (well B in Figure 14.6). This is known as an **artesian well**, because the water rises above the top of the aquifer. In some situations, the potentiometric surface may be above the ground level. The water in a well drilled into the confined aquifer in this situation would rise above ground level, and flow out, if it's not capped (well C in Figure 14.6). This is known as a **flowing artesian well**.

Figure 14.6 A depiction of the water table and the potentiometric surface of a confined aquifer. [SE]

In situations where there is an aquitard of limited extent, it is possible for a perched aquifer to exist as shown in Figure 14.7. Although perched aquifers may be good water sources at some times of the year, they tend to be relatively thin and small, and so can easily be depleted with over-pumping.

Figure 14.7 A perched aquifer above a regular unconfined aquifer. [SE]

In 1856, French engineer Henri Darcy carried out some experiments from which he derived a method for estimating the rate of groundwater flow based on the hydraulic gradient and the permeability of an aquifer, expressed using *K*, the hydraulic conductivity. Darcy's equation, which has been used widely by hydrogeologists ever since, looks like this:

$V = K * i$

(where V is the velocity of the groundwater flow, K is the hydraulic conductivity, and i is the hydraulic gradient).

We can apply this equation to the scenario in Figure 14.5. If we assume that the permeability is 0.00001 m/s we get: $V = 0.00001 * 0.08 = 0.0000008$ m/s. That is equivalent to 0.000048 m/min, 0.0029 m/hour or 0.069 m/day. That means it would take 1,450 days (nearly four years) for water to travel the 100 m from the vicinity of the well to the stream. Groundwater moves slowly, and that is a reasonable amount of time for water to move that distance. In fact it would likely take longer than that, because it doesn't travel in a straight line.

Exercises

Exercise 14.1 How Long Will It Take?

Using $V = K * i$, estimate the velocity of groundwater flow from Joe's to the stream, and determine how long it might take for contaminated groundwater to flow the 80 m to the stream. [SE drawing]

Sue, the owner of Joe's 24-Hour Gas, has discovered that her underground storage tank (UST) is leaking fuel. She calls in a hydrogeologist to find out how long it might take for the fuel contamination to reach the nearest stream. They discover that the well at Joe's has a water level that is 37 m above sea level and the elevation of the stream is 21 m above sea level. The sandy sediment in this area has a permeability of 0.0002 m/s.

It's critical to understand that groundwater does not flow in underground streams, nor does it form underground lakes. With the exception of **karst** areas, with caves in limestone, groundwater flows very slowly through granular sediments, or through solid rock that has fractures in it. Flow velocities of several centimetres per day are possible in significantly permeable sediments with significant hydraulic gradients. But in many cases, permeabilities are lower than the ones we've used as examples here, and in many areas, gradients are much lower. It is not uncommon for groundwater to flow at velocities of a few millimetres to a few centimetres per year.

As already noted, groundwater does not flow in straight lines. It flows from areas of higher hydraulic head to areas of lower hydraulic head, and this means that it can flow "uphill" in many situations. This is illustrated in Figure 14.8. The dashed orange lines are **equipotential,** meaning lines of equal pressure. The blue lines are the predicted groundwater **flow paths**. The dashed lines red lines are no-flow boundaries, meaning that water cannot flow across these lines. That's not because there is something there to stop it, but because there's no pressure gradient that will cause water to flow in that direction.

Groundwater flows at right angles to the equipotential lines in the same way that water flowing down a slope would flow at right angles to the contour lines. The stream in this scenario is the location with the lowest hydraulic potential, so the groundwater that flows to the lower parts of the aquifer has to flow upward to reach this location. It is forced upward by the pressure differences, for example, the difference between the 112 and 110 equipotential lines.

Figure 14.8 Predicted equipotential lines (orange) and groundwater flow paths (blue) in an unconfined aquifer. The orange numbers are the elevations of the water table at the locations shown, and therefore they represent the pressure along the equipotential lines. [SE]

Groundwater that flows through caves, including those in **karst** areas — where caves have been formed in limestone because of dissolution — behaves differently from groundwater in other situations. Caves above the water table are air-filled conduits, and the water that flows within these conduits is not under pressure; it responds only to gravity. In other words, it flows downhill along the gradient of the cave floor (Figure 14.9). Many limestone caves also extend below the water table and into the saturated zone. Here water behaves in a similar way to any other groundwater, and it flows according to the hydraulic gradient and Darcy's law.

Figure 14.9 Groundwater in a limestone karst region. The water in the caves above the water table does not behave like true groundwater because its flow is not controlled by water pressure, only by gravity. The water below the water table does behave like true groundwater. [SE]

14.3 Groundwater Extraction

Except in areas where groundwater comes naturally to the surface at a **spring** (a place where the water table intersects the ground surface), we have to construct wells in order to extract it. If the water table is relatively close to the surface, a well can be dug by hand or with an excavator, but in most cases we need to use a drill to go down deep enough. There are many types of drills that can be used; an example is shown in Figure 14.10. A well has to be drilled at least as deep as the water table, but in fact must go much deeper; first, because the water table may change from season to season and from year to year, and second, because when water is being pumped, the water level will drop, at least temporarily.

Where a well is drilled in unconsolidated sediments or relatively weak rock, it has to be lined with casing (steel pipe in most cases) in order to ensure that it doesn't cave in. A specially designed well screen is installed at the bottom of the casing. The size of the holes in the screen is carefully chosen to make sure that it allows the water to move into the well freely, but prevents aquifer particles from entering the well. A submersible pump is typically used to lift water from within the well up to the where it is needed. The well shown in Figure 14.10 has casing that is about 40 cm in diameter, which might be typical for a municipal water supply well, or a very large well for irrigation. Most domestic wells have 15 cm casing.

Figure 14.10 A water-well drilling rig in operation in the Cassidy area, near Nanaimo, B.C. In the photo on the right the well is being test-pumped with air pressure. The casing (yellow arrow) is about 40 cm in diameter. [SE]

Pumping water from the well removes water from inside the well at first. That lowers the water level inside the well. This means that water will flow from the surrounding aquifer (higher groundwater head) toward the pumping well where the groundwater head is now lower. That is how a well gets water from the ground. The water table, or potentiometric surface, will slope in toward the well where the water is being withdrawn. That indicates the energy gradient that is allowing water to flow toward the well. This creates a shape known as a **cone of depression** surrounding the well, as illustrated in Figure 14.11. If pumping from a well continues for hours to days, the cone of depression may result in a loss of water in nearby wells. As shown in Figure

14.12, pumping of well C has contributed to well B going dry. If pumping continues in well C, it too may go dry.

Figure 14.11 Three wells in an unconfined aquifer. Well A is not being pumped. Well B is being pumped at a slow rate and well C, which has a larger cone of depression, is being pumped at a faster rate. [SE]

Figure 14.12 A similar scenario to that in Figure 14.11, but in this case, wells B and C have been pumped unsustainably for a long time. The cone of depression from well C has reached well B and has contributed to it going dry. [SE]

Exercises

Exercise 14.2 Cone of Depression

How will this likely affect the rate of flow into the well?

The two diagrams here show the same well before (left) and after (right) long-term pumping. A cone of depression has developed. This provides the energy gradient for water to flow toward the well so that it can be pumped out.

Like other provinces in Canada, British Columbia has a network of observation wells administered by the Ministry of the Environment. These are wells that are installed to measure water levels; they are not pumped. There are 145 active observation wells in B.C. (in 2015), most equipped with automatic recorders that monitor water levels continuously. The main purpose of the observation wells is to monitor water table levels so that we can see if there are long-term natural fluctuations in groundwater quantity, and shorter-term fluctuations related to overuse of the resource. They are also sampled regularly to monitor groundwater chemistry and quality.

An example of an observation well is illustrated in Figure 14.13. This one is situated at Cassidy on Vancouver Island and is used to monitor an unconsolidated aquifer that is widely used by residents with private wells.

Figure 14.13 B.C. observation well 232 near Cassidy Airport, Vancouver Island. The installation also has a solar panel, which is not visible in this view. [SE]

The water-level data from B.C.'s observation wells are available to the public, and an example data set is illustrated in Figure 14.14. The water level in Ministry of Environment observation well 232 (OW-232), situated in Lantzville on Vancouver Island, dropped significantly from 1979 (average depth ~1.5 m), to 2010 (average depth ~5.5 m), but has recovered a little since then.

Figure 14.14 Water level data for B.C. observation well 232 on Harby Rd., Lantzville, Vancouver Island. From 1979 to 2003, depths were recorded monthly. Automated equipment was installed in 2003, and the depths were recorded hourly since that time. [SE from data at: http://www.env.gov.bc.ca/wsd/data_searches/obswell/map/]

14.4 Groundwater Quality

As was noted at the very beginning of this chapter, one of the good things about groundwater as a source of water is that it is not as easily contaminated as surface water is. But there are two caveats to that: one is that groundwater can become naturally contaminated because of its very close connection to the materials of its aquifer, and the second is that once contaminated by human activities, groundwater is very difficult to clean up.

Natural Contamination of Groundwater

Groundwater moves slowly through an aquifer, and unlike the surface water of a stream, it has a lot of contact with the surrounding rock or sediment. In most aquifers, the geological materials that make up the aquifer are relatively inert, or are made up of minerals that dissolve very slowly into the groundwater. Over time, however, all groundwater gradually has more and more material dissolved within it as it remains in contact with the aquifer. In some areas, that rock or sediment includes some minerals that could potentially contaminate the water with elements that might make the water less than ideal for human consumption or agricultural use. Examples include copper, arsenic, mercury, fluorine, sodium, and boron. In some cases, contamination may occur because the aquifer material has particularly high levels of the element in question. In other cases, the aquifer material is just normal rock or sediment, but some particular feature of the water or the aquifer allows the contaminant to build up to significant levels.

An example of natural contamination takes place in the bedrock aquifers of the east coast of Vancouver Island and the adjacent Gulf Islands. The aquifer is the Cretaceous (90 Ma to 65 Ma) Nanaimo Group, which is made up of sandstone, mudstone, and conglomerate (Figure 14.17).

Figure 14.17 Cretaceous Nanaimo Group sandstone exposed in a Nanaimo parking lot [SE]

The rocks of the Nanaimo Group are not particularly enriched in any trace elements, but the submarine-fan sandstone that makes up much of the group is a lithic wacke, and therefore has relatively high levels of clay (for a sandstone). This clay is good at adsorbing[1] some elements from the water and desorbing others, and in the process, its pH goes up (it becomes alkaline). At high pH levels (some as high as 9 in the Nanaimo Group), the element fluorine that is present

naturally in the rock (as it is in almost any rock) has an increased tendency to dissolve in the water. In some areas, groundwater in the Nanaimo Group has fluorine levels that are well above recommended levels for drinking water. The World Health Organization (WHO) maximum acceptable concentration (MAC) for fluorine is 1.5 mg/L (milligrams per litre). Between 5% and 10% of the domestic wells around Nanaimo and adjacent Gabriola Island have more than that, some as much as 10 mg/L. A small amount of fluorine in the human diet is considered important for maintaining dental health, but high levels can lead to malformation and discolouration of teeth, and long-term exposure can lead to other more serious health effects such as skeletal problems.

Nanaimo Group groundwater can also have elevated levels of boron, again related to pH and adsorption from clay minerals. While boron at the levels found there is not toxic to humans, there is enough boron in some wells to be toxic to plants, and the water cannot be used for irrigation.

Rural residents in the densely populated country of Bangladesh (over 1,000 residents/km2, compared with 3.4/km2 in Canada) used to rely mostly on surface supplies for their drinking water, and many of these were subject to bacterial contamination. Infant mortality rates were among the highest in the world and other illnesses such as diarrhea, dysentery, typhoid, cholera, and hepatitis were common. In the 1970s, international agencies, including UNICEF, started a program of drilling wells to access abundant groundwater supplies at depths of 20 m to 100 m. Eventually over 8 million such wells were drilled. Infant mortality and illness rates dropped dramatically, but it was later discovered that the water from a high proportion of these wells has arsenic above safe levels (Figure 14.18).

Figure 14.18 The distribution of arsenic in groundwater in Bangladesh. The WHO recommended safe level for arsenic is 10 μg/L. All of the green, orange, and red areas on the map exceed that limit.
[From: BGS and DPHE. 2001. Arsenic contamination of groundwater in Bangladesh. Kinniburgh, D G and Smedley, P L (Editors). British Geological Survey Technical Report WC/00/19. British Geological Survey: Keyworth.
(http://www.bgs.ac.uk/arsenic/bangladesh/)]

Most of the wells in the affected areas are drilled into relatively recent sediments of the vast delta of the Ganges and Brahmaputra Rivers. While these sediments are not particularly enriched in arsenic, they have enough organic matter in them to use up any oxygen present. This leads to water with a naturally low oxidation potential (anoxic conditions); arsenic is highly soluble under these conditions, and so any arsenic present in the sediments easily gets dissolved into the groundwater. Arsenic poisoning leads to headaches, confusion, and diarrhea, and eventually to vomiting, stomach pain, and convulsions. If not treated, the final outcomes are heart disease, stroke, cancer, diabetes, coma, and death. There are ways to treat arsenic-rich groundwater, but it is a challenge in Bangladesh to implement the simple and effective technology that is available.

Anthropogenic Contamination of Groundwater

Groundwater can become contaminated by pollution at the surface (or at depth), and there are many different anthropogenic (human-caused) sources of contamination.

The vulnerability of aquifers to pollution depends on several factors, including the depth to the water table, the permeability of the material between the surface and the aquifer, the permeability of the aquifer, the slope of the surface, and the amount of precipitation. Confined aquifers tend to be much less vulnerable than unconfined ones, and deeper aquifers are less vulnerable than shallow ones. Steeper slopes mean that surface water tends to run off rather than infiltrate (and this can reduce the possibility of contamination). Contamination risk is also less in dry areas than in areas with heavy rainfall.

Studies of groundwater vulnerability have been completed for various regions of British Columbia. A groundwater vulnerability map for southern Vancouver Island is shown in Figure 14.19. The yellow to red areas are considered to have high vulnerability to pollution from surface sources, and most of these are where the aquifers are unconfined in quite permeable unconsolidated sediments of either glacial or fluvial origin, where the water table is relatively shallow and the terrain is relatively flat.

Figure 14.19 The vulnerability to anthropogenic contamination of aquifers on southern Vancouver Island. Much of the island is not mapped (shown as white) because of a lack of aquifer information in areas without wells. [From: Newton, P. and Gilchrist, A. 2010. Technical summary of intrinsic vulnerability mapping methods of Vancouver Island, Vancouver Island Water Resources Vulnerability Mapping Project, Vancouver Island University, 45pp. Used with permission. https://web.viu.ca/groundwater/PDF/VI_DRASTIC_Summary_Phase2_2010.pdf]

The important sources of anthropogenic groundwater contamination include the following:

- Chemicals and animal waste related to agriculture, and chemicals applied to golf courses and domestic gardens
- Landfills
- Industrial operations
- Mines, quarries, and other rock excavations
- Leaking fuel storage tanks (especially those at gas stations)
- Septic systems
- Runoff from roads (e.g., winter salting) or chemical spills of materials being transported

Agriculture

Intensive agricultural operations and golf courses can have a significant impact on the environment, especially where chemicals and other materials are used to enhance growth or control pests. An example of agricultural contamination is in the Abbotsford area of the Fraser Valley, where nitrate levels above the 44 mg/L maximum acceptable level (expressed as nitrate) in the Abbotsford-Sumas aquifer have been observed since the 1950s; however, the problem became much worse as agriculture intensity increased in the 1980s. By 2004, groundwater with nitrate levels in excess of 44 mg/L was reported over an area of about 75 km2 around Abbotsford, and the problem extended across the border into the Sumas area of Washington State.

This region is intensively used for berry crops (especially raspberries and blueberries) and large poultry operations, as well as lesser amounts of grazing and forage crops. Chicken manure is typically stored in fields adjacent to chicken barns, and may release nitrogen to the environment from runoff water, and from releases of ammonia gas. Over decades, both chemical fertilizers and chicken manure and other manures have been applied to the berry crops to provide extra nitrogen to help maximize berry growth. If the fertilizer added is in excess of what the plants need, or is poorly timed compared to when it is needed, then the extra nitrogen may be leached into the groundwater below. Berry crops are irrigated over the summer to help the crops grow. Summer irrigation and winter rainfall may carry excess nitrate from the near surface to the aquifer below.

Since the 1990s, agricultural practices have been tightened up to reduce the rate of groundwater contamination, but it will take decades for nitrate levels to drop in the Abbotsford-Sumas aquifer. Agriculture and Agri-Food Canada and many others are conducting research on better irrigation and nitrate management techniques to reduce the amount of nitrogen that leaches to groundwater.

Landfills

In the past, domestic and commercial refuse was commonly trucked to a “dump” (typically a hole in the ground), and when the hole was filled, it was covered with soil and forgotten. In situations like this, rain and melting snow can easily pass through the soil used to cover the refuse. This water passes into the waste itself, and the resulting landfill **leachate** that flows from the bottom of the landfill can seriously contaminate the surrounding groundwater and surface water. In the past few decades, regulations around refuse disposal have been significantly

strengthened, and important steps have been taken to reduce the amount of landfill waste by diverting recyclable and compostable materials to other locations.

A modern engineered landfill has an impermeable liner (typically heavy plastic, although engineered clay liners or natural clay may be adequate in some cases), a plumbing system for draining leachate (the rainwater that flows through the refuse and becomes contaminated), and a network of monitoring wells both within and around the landfill (Figure 14.20). Once part or all of a landfill site is full, it is sealed over with a plastic cover, and a system is put in place to extract **landfill gas** (typically a mixture of carbon dioxide and methane). That gas can be sent to a nearby location where it is burned to create heat or used to generate electricity. The leachate must be treated, and that can be done in a normal sewage treatment plant.

Figure 14.20 A cross-section of a typical modern landfill [SE]

The monitoring wells are used to assess the level of the water table around the landfill and to collect groundwater samples so that any leakage can be detected. Because some leakage is almost inevitable, the ideal placement for landfills is in areas where the depth to the water table is significant (tens of metres if possible) and where the aquifer material is relatively impermeable. Landfills should also be situated far from streams, lakes, or wetlands so that contamination of aquatic habitats can be avoided.

Today there are hundreds of abandoned dumps scattered across the country; most have been left to contaminate groundwater that we might wish to use sometime in the future. In many cases, it's unlikely that we'll be able to do so.

Exercises

Exercise 14.4 What Goes on at Your Landfill?

Unless you live in a remote rural area, there's a good chance that the refuse you can't recycle is picked up at the curb and taken to a landfill. Most landfills are operated by cities or regional districts, and you should be able to find information about yours on the appropriate local

government website. See if you can answer some the following questions:

1. Which government body operates your landfill?

2. Where is the landfill situated?

3. Is your waste all placed in a landfill, or are there other processes in use (e.g., incineration or composting)?

4. Are landfill gases captured, and, if so, what is done with them?

5. What could be changed to improve the waste disposal situation in your community (e.g., more recycling, compost collection, waste-to-energy technology)?

[SE photo]

Industrial Operations

Although western Canada doesn't have the same extent of industrial pollution as other parts of the country, there are still seriously contaminated sites in the west, most with the potential to contaminate groundwater. One example is the lead and zinc smelter at Trail, B.C. The largest in the world, it has been operating for over 100 years and has left a residue of metal contamination around the region (Figure 14.21). In some parts of Trail, the contamination is serious enough that existing soil has been removed from residential properties and replaced with clean soil brought in from elsewhere. This contaminated soil has contributed to contamination of groundwater in the Trail area. Groundwater beneath the actual smelter site is contaminated, and the operator (Teck Resources) is currently working on plans to prevent that water from reaching the nearby Columbia River.

Figure 14.21 The Trail lead-zinc smelter in 1929
[http://upload.wikimedia.org /wikipedia/commons/2/20 /Trail_Smelter_in_Year_1929.png]

Mines, Quarries, and Rock Excavations

Mines and other operations that involve the excavation of large amounts of rock (e.g., highway construction) have the potential to create serious environmental damage. The exposure of rock that has previously not been exposed to air and water can lead to the oxidation of sulphide-bearing minerals, such a pyrite (FeS2), within the rock. The combination of pyrite, water, oxygen, and a special type of bacteria (*Acidithiobacillus ferrooxidans*) that thrives in acidic conditions leads to the generation of acidity, in some cases to pH less than 2. Water that acidic is hazardous by itself, but the low pH also has the property of increasing the solubility of certain heavy metals. The water that is generated by this process is known as acid rock drainage (ARD). ARD can occur naturally where sulphide-bearing rocks are near the surface. The issue of ARD is a major environmental concern at both operating mines and abandoned mines (see Chapter 20). In streams around the Mt. Washington Mine on Vancouver Island (Figure 14.22), copper levels are high enough to be toxic to fish. Groundwater adjacent to the contaminated streams in the area is very likely contaminated as well.

Figure 14.22 Acidic runoff at the abandoned Mt. Washington Mine near Courtenay, B.C. [SE]

Leaking Fuel Tanks

Underground storage tanks (USTs) are used to store fuel at gas stations, industrial sites, airports, and anywhere that large volumes of fuel are used. They do not last forever, and eventually they start to leak their contents into the ground. This is a particular problem at older gas stations — although it may also become a future problem at newer gas stations. You may have noticed gas stations that have been closed and then surrounded by chain-link fence (Figure 14.23). In virtually all such cases the closure has been triggered by the discovery of leaking USTs and the requirement to cease operations and remediate the site.

Figure 14.23 A closed and fenced gas station site in Nanaimo, B.C. The white pipes in the background are wells for monitoring groundwater contamination on the site. [SE]

Petroleum fuels are complex mixtures of hydrocarbon compounds and the properties of their components — such as density, viscosity, solubility in water, and volatility — tend to vary widely. As a result, a petroleum spill is like several spills for the price of one. The petroleum liquid slowly settles through the unsaturated zone and then tends to float on the surface of the groundwater (Figure 14.24). The more readily soluble components of the spill dissolve in the groundwater and are dispersed along with the normal groundwater flow, and the more volatile components of the spill rise toward the surface, potentially contaminating buildings.

Figure 14.24 A depiction of the fate of different components of a petroleum spill from an underground storage tank. [SE]

Exercises

Exercise 14.5 Find a Leaking UST in Your Community

There is almost certainly a leaking UST at a former gas station near you. Look for an empty property that is surrounded by a chain-link fence with “No Trespassing” signs. You might see evidence of monitoring wells (like those shown in Figure 14.24), and there could be some petroleum barrels around that are being used to store contaminated water. Once you’ve identified one of these, you’ll probably start seeing them everywhere!

Septic Systems

In areas that are not served by sewage networks leading to a central sewage treatment plant, most homeowners rely on **septic systems** for disposal of sewage. There are two primary components to a simple septic system, the septic tank and the drainage field (Figure 14.25). A typical septic tank is constructed of either concrete or plastic and has a volume of 5,000 L to 10,000 L (5 m3 to 10 m3). This forms the first treatment and is designed to be **anaerobic** (without oxygen). That promotes the activity of certain bacteria that help break down the waste. As the waste is degraded, some portions tend to sink to form sludge at the base of the tank, and others float to the surface, forming a scum layer. A septic tank may be divided into two parts to keep the sludge at the bottom and the scum on the top from draining out. The water then moves to the drainage field, which provides the right conditions for a different set of bacteria that operate in **aerobic** conditions. The drainage field includes an array of plastic pipes that are perforated to allow the effluent to drain out over a large area and seep slowly into the ground. In order to install a drainage field, it is first necessary to test the soil below, as it must be sufficiently permeable to allow the effluent to percolate away, but not so permeable that it flows too quickly and the soil is not able to filter out the pathogenic bacteria.

If they are properly installed and used, and if the sludge is periodically removed from the tank, a septic system should be effective in treating the sewage for decades. The anaerobic and aerobic bacteria should be able to break down the incoming waste and there should be little risk to the surface environment or groundwater. But many things can go wrong with a septic system, including the following:

- If inappropriate chemicals are added to the waste stream, they may interfere with the natural breakdown of the sewage.
- If the tank is not periodically pumped out, solids can get into the drainage field and compromise the drainage, resulting in the flow of effluent toward the surface.
- If the soil is either not sufficiently permeable or too permeable, the effluent will not drain away (and will start to pool at the surface) or it will drain too quickly.
- If the drainage field is constructed in an area where the water table is close to surface, some of the effluent is likely to flow into the groundwater without being treated.

Figure 14.25 A typical septic system. [SE]

Prevention and Mitigation of Groundwater Contamination

As illustrated in the landfill example above, there are two fairly simple ways to significantly reduce the chance and degree of groundwater contamination from surface sources. One is to prevent rainwater from infiltrating down to the water table and picking up contaminants; this can be achieved by simply capping or roofing over the landfill, mine tailings, or spill site. The second is to provide an impermeable barrier beneath the contaminant. Modern landfills and mine tailings impoundments are all built using some combination of clay and engineered plastic barriers. Both of these solutions — caps and liners — are subject to failure due to leaks.

Once contaminants are in the groundwater, the main form of remediation is to pump out the contaminated water and treat it at the surface. This can be a slow process, and preventing the contaminant from travelling significantly during this process can be accomplished by manipulating local groundwater flow through the extraction or injection of water at certain locations. Consider this in the exercise below.

Exercises

Exercise 14.6 Manipulating a Contaminant Plume

This diagram shows a groundwater contaminant plume in red. The source of the contamination has been removed but if the plume is not dealt with, it will eventually enter the stream and threaten the health of wildlife. Pumping the contaminant from well B for treatment will not be sufficient to prevent some of the contamination from making it to the stream.

What could you do at wells A and C to prevent this? Explain and use the diagram below to illustrate the expected changes to the water table and the movement of the plume.

---

1. "Adsorb" (with a "d") is not the same as "absorb" (with a "b"). Water can be absorbed by a sponge. Ions dissolved in water can be adsorbed onto — or desorbed from — the surfaces of clay minerals. ↲

Chapter 14 Summary

The topics covered in this chapter can be summarized as follows:

14.1	Groundwater and Aquifers	Porosity is the percentage of open space within a rock or unconsolidated sedimentary deposit, while permeability is the facility with which water can be transmitted through that material. An aquifer is a body of rock or sediment that has sufficient permeability for water to be extracted, while an aquitard is an impermeable body. An aquifer is described as confined if it is overlain by an impermeable layer (confining layer), or unconfined if it has no such confining layer.
14.2	Groundwater Flow	The water table is the upper surface of the saturated zone in an unconfined aquifer. A confined aquifer has a potentiometric surface (instead of a water table), which is defined as the level to which water would rise if a well were drilled into the confined aquifer. Change in groundwater head over distance is the hydraulic gradient. The theoretical velocity of flow in an aquifer is defined by Darcy's law as the hydraulic conductivity (a measure of permeability) times the hydraulic gradient ($V = K * i$). It is possible to predict groundwater flow paths if we can draw equipotential lines within an aquifer. In areas where limestone has solutional openings (e.g., caves), water flow is determined by gravity above the water table and by the hydraulic gradient below the water table.
14.3	Groundwater Extraction	Groundwater can be extracted at springs, but in most cases, wells are needed to ensure a steady supply. Pumping groundwater from a well lowers groundwater head near the well, creating flow toward the well. This creates a cone of depression around the well. Excessive pumping can lead to a well running dry or

		to a lack of water in nearby wells. During extended periods of dry weather, or if consistent over-pumping occurs, aquifers may be depleted. Observation wells are used to monitor short-term and long-term changes in water levels that can indicate changes in aquifer health.
14.4	Groundwater Quality	The quality of groundwater can be compromised by both natural and anthropogenic contamination. Natural contamination can be caused by particularly high levels of contaminants within the aquifer itself, but is more commonly a result of enhanced solubility of contaminants due to the aquifer chemistry. Some common sources of anthropogenic contamination include agriculture, industry, mining, landfills, and leaking underground storage tanks. We can assess the vulnerability of aquifers to contamination by mapping regional variations in parameters such as depth to the water table, permeability, slope, and precipitation.

Questions for Review

1. What is the difference between porosity and permeability?2. Both sand and clay deposits can have high porosity, but while most sand also has high permeability, clay does not. Why not?

3. Arrange the following types of rock in order of their likely permeability, as measured by the hydraulic conductivity (*K*): mudstone, fractured granite, limestone in a karst region, sandstone, and unfractured gneiss.

4. Sue, the owner of Joe's 24-Hour Gas, has a shallow well (15 m deep) as illustrated in the diagram. The well can only produce 0.5 L per minute, but that's enough for water to make coffee and supply a washroom that gets used several times a day. Frank, who operates a raspberry farm next door, uses up to 250,000 L of water per day to irrigate his crop during summer. He gets water from a deeper well that can produce 250 L/minute. See the diagram below. (a) What type of aquifer does Sue use? (b) What type of aquifer does Frank use? (c) It seems that what Sue calls an aquifer is an aquitard (confining layer) from Frank's perspective. How is that possible?

5. Two wells 70 m apart have water levels of 77 m and 83 m above sea level respectively. The aquifer has a hydraulic conductivity of 0.0003 m/s. What is the likely velocity of groundwater flow in the region between these two wells?

6. The well in question 5 with a water level of 83 m is heavily used and after several months the water level has dropped by 9 m. How will that affect the flow of groundwater in the area between the two wells?

7. Explain why it is important for provincial governments to operate observation well networks.

8. What is the main difference between natural and anthropogenic contamination of groundwater?

9. Why is a highly permeable aquifer more vulnerable to anthropogenic contamination than a less permeable aquifer?

10. How can a livestock operation lead to contamination of groundwater? What is the most likely contaminant?

11. Which mineral in the rock of a mining operation is typically responsible for acid rock drainage?

12. Why is it necessary to test the permeability of the soil before constructing a septic field?

Chapter 21 Geological History of Western Canada

Introduction

Learning Objectives

After reading this chapter, completing the exercises within it, and answering the questions at the end, you should be able to:

- Describe the general makeup and ages of the provinces of Laurentia, Laurentia's journey over the past 650 Ma, and the processes by which additional rocks were added on its eastern, northern, and western margins during the Phanerozoic to form the continent of North America
- Explain the timing and depositional environments of mid- and late-Proterozoic sedimentary rocks in western Canada
- Describe the depositional environments and types of sedimentary rock that accumulated on the western margin of North America and in the Western Canada Sedimentary Basin (WCSB) during the Paleozoic
- Summarize the extents, geological origins, and migration of the accreted terranes of British Columbia and Yukon
- Explain how terrane accretion on the west coast during the Mesozoic contributed to the formation of the Rocky Mountains and how that in turn provided the source material for a thick sequence of Mesozoic sedimentary rock in the WCSB
- Describe the origins of the Mesozoic intrusive igneous rocks of the Coast Range and other areas within British Columbia
- Describe the geological effects of the accretion of the Pacific Rim and Crescent Terranes, the nature of WCSB deposition in the early Cenozoic, the ongoing volcanism and earthquake activity in western Canada, and the general effects of the Pleistocene glaciation in western Canada

Figure 21.1 Crowsnest Mountain in the southern Alberta Rockies is made up of Paleozoic rocks that were uplifted by continental convergence during the Mesozoic, and then eroded by glaciation during the Cenozoic [SE]

Western Canada has a fascinating geological history with rocks ranging in age from the Archean to the Holocene. Over that time, almost every conceivable geological process has taken place here, resulting in the formation of a wide array of rock types, and some of the most important fossil deposits in the world. The region is also endowed with a range of geological resources, spanning the periodic table from beryllium to uranium, and the geological processes have produced awe-inspiring scenery and world-class recreational opportunities.

This chapter focuses on the important geological history and geological features of western Canada, but includes an overview of Canadian geology as a whole, starting with the development and journey of the ancient continent of Laurentia.

21.1 Geological History of Canada

Laurentia, which makes up the core of North America, is the largest and arguably the oldest of Earth's **cratons** (regions of stable ancient crust). Some of the rocks are over 4 billion years old, and Laurentia has been together in its present form for the last billion years. Over the past 650 million years, Laurentia has moved along a zigzag path from deep in the southern hemisphere to close to the North Pole (Figure 21.2). During that time, it collided several times with other continents and was temporarily part of two supercontinents (Pannotia and Pangea).

Figure 21.2 The path of Laurentia over the past 650 Ma [SE]

Bodies of rock tend to be eroded and recycled through the processes of plate tectonics, including uplift leading to erosion and burial leading to melting, and thus there are very few areas of truly ancient rocks on Earth. The oldest undisputed rocks are those of the Acasta Gneiss from north of Yellowknife, Northwest Territories, aged 4.03 Ga. But there are some rocks that could be even older within the Nuvvuagittuq greenstone belt on the east coast of Hudson Bay, in Quebec. These have been isotopically dated at 4.28 Ga, although the reliability of that date has been questioned. Based on other data, it is acknowledged that the Nuvvuagittuq rocks are at least as old as 3.75 Ga. The Acasta and Nuvvuagittuq rocks are situated within the Slave and Superior Cratons respectively, the oldest parts of Laurentia (Figure 21.3). Although these ancient cratons are not consistently that old, they are generally older than 3 Ga, as is part of the Wyoming Craton. The Hearne and Rae Cratons are older than 2 Ga, while most of the other parts of Laurentia are aged between 1 Ga and 2 Ga. The various provinces of Laurentia were assembled by plate-tectonic processes between 1 Ga and 3 Ga.

The areas of Figure 21.3 that are left uncoloured — the Appalachian, Innuitian, and Cordilleran fold belts — are

geological regions that have been added to North America since 500 Ma. These are at least partly made up of sedimentary rocks that were deposited along the coasts and then folded, faulted, and uplifted during continental collisions.

The term *Laurentia* is geologically equivalent to the term **Canadian Shield**, although the latter is generally considered to be the area where the ancient Laurentian rocks are exposed at the surface and not covered with younger rocks. That applies to most of the region to the north and east of the red dotted line in Figure 21.3.

Figure 21.3 The main provinces of Laurentia. The pink areas are the oldest; light yellow are the youngest. All of the areas south and west of the dotted red line are now covered with younger rocks. The white areas represent rocks that were added to North America since 700 Ma. [SE]

Laurentia was part of the supercontinent Rodinia during the period between 1,100 Ma and 700 Ma. As Rodinia started to break up after 700 Ma, sediments derived from the erosion of the interior of the continent began to accumulate along its coasts, initially along the west coast, then the east coast at around 600 Ma, and finally on the north coast by around 550 Ma. This process continued for several hundred million years. By around 450 Ma, large areas of the interior of Laurentia were depressed below sea level — probably because of the downward pull of an underlying subducting plate — and marine sediments were deposited over parts of Quebec, Ontario, Manitoba, Saskatchewan, Alberta, and the Northwest Territories during the Ordovician, Silurian, and Devonian Periods (450 Ma to 350 Ma). These sediments are coloured various shades of blue on the geological map of Canada (Figure 21.4).

Figure 21.4 Geological map of Canada from the Geological Survey of Canada [http://geoscan.nrcan.gc.ca/starweb/geoscan/servlet.starweb?path=geoscan/fulle.web&search1=R=208175]

At approximately 350 Ma, the part of Gondwana that is now Africa collided with the eastern coast of North America, thrusting volcanic islands and sedimentary layers far inland to become the Appalachian fold belt. The Appalachian Mountains would have rivalled the Himalayas in extent and height during the Devonian. At about the same time, a smaller continent, Pearya, collided with the north coast, creating the Innuitian fold belt.

At around 200 Ma, small continents that now make up the interior of B.C. and part of Yukon collided with the west coast of North America, starting the process of thrusting the sedimentary rocks inland and upward to form the Rocky Mountains.

The west-central part of North America subsided once again at around 150 Ma, due to an underlying subducting plate, and this led to the deposition of more marine rocks across Manitoba, Saskatchewan, and Alberta, and north into the Northwest Territories and Yukon (the green areas in Figure 21.3).

Finally, at around 90 Ma, more small continents, which now comprise Vancouver Island and Haida Gwaii, collided with the west coast, leading to further uplift of the Rocky Mountains.

Exercises

Exercise 21.1 Finding the Geological Provinces of Canada

Figure 21.4 shows the geology of Canada in some detail, with the colours representing the lithologies and ages of the rocks. Identifying in Figure 21.4 some of the major features shown in Figure 21.3 will help you understand the distribution of Canada's geological features. Start by outlining the extent of the exposed Canadian Shield (the dotted red line); you might also be able to identify some of the cratons within the Shield. Then look for the limits of the Appalachian and Innuitian fold belts. Finally pick out the extent of the Cordilleran fold belt.

The best way to do this would be to print out a copy of Figure 21.4 and draw the boundaries from Figure 21.3 on it with a pencil. If you're interested, you can get your own high resolution copy of the geological map of Canada at http://geoscan.nrcan.gc.ca/starweb/geoscan/servlet.starweb?path=geoscan/fulle.web&search1=R=208175.

21.2 Western Canada during the Precambrian

Laurentia extends as far west as eastern B.C. (Figure 21.3), but the ancient rocks of the craton are almost completely covered by younger rocks in B.C., Yukon, and all of Alberta except the far northeast corner. Laurentia is well represented in northern Saskatchewan and across large parts of Manitoba, the Northwest Territories, and Nunavut (Figure 21.5). Where they are exposed, the rocks of the Canadian Shield are highly varied lithologically, typically strongly metamorphosed due to their deep burial at some time in the past, and in some cases, quite different from what could be expected to occur on Earth today.

Starting from the south, in eastern Manitoba and adjacent Ontario, we have the ancient rocks of the Superior Province. On the map the Superior Province, rocks are mostly pink, representing granitic and gneissic rocks, with strips and blotches of green, representing metamorphosed sea-floor basalt and sediments, also known as greenstone belts. These rocks are widely interpreted to have deep crustal origins, and include large areas of granulite facies metamorphic rock formed at high temperatures and moderate to high pressures (see Figure 7.19). Superior Province greenstone belts in Ontario and Quebec host some of the world's largest volcanogenic massive sulphide deposits. As described in Chapter 20, the Superior Province in northern Manitoba is host to important nickel deposits at Thompson. These formed from mantle-derived mafic magma that interacted with sulphur-bearing crustal rocks, and within which heavy-metal sulphide minerals formed.

The Trans-Hudson Orogen (THO), as its name implies, extends through Saskatchewan and Manitoba and over to the eastern side of Hudson Bay. It represents the continent-continent collision zone between the Superior Craton to the south and the Churchill Craton (including the Wyoming, Hearne, and Rae Cratons) to the north; thus it's a remnant of the initial formation of Laurentia at around 1.9 Ga. At the time of the collision, the THO would have been a major mountain range, and the rocks that we see there now — which evolved deep beneath those mountains — are highly metamorphosed sedimentary and volcanic rocks intruded by large granitic bodies. The important volcanogenic massive sulphide deposits around Flin Flon are within the THO.

Figure 21.5 Geological features of the Canadian Shield of western Canada. A.B.: Athabasca Basin, T.B.: Thelon Basin, and TMZ: Taltson Magmatic Zone [By SE after: http://geoscan.nrcan.gc.ca/starweb/geoscan/servlet.starweb?path=geoscan/fulle.web&search1=R=208175]

The Churchill Craton is lithologically similar to the Superior Craton, although not generally as old. It includes two important sedimentary basins: the Athabasca Basin in Saskatchewan and the Thelon Basin in Nunavut, both filled with rocks aged around 1.7 Ga. These consist primarily of sandstones and minor mudstones that are only weakly metamorphosed and essentially undeformed (not folded) because they are situated within a stable craton and so have not been subjected to significant tectonic forces. The Athabasca Basin is economically important for its large and rich unconformity-type uranium deposits (see Chapter 20). At its western end, there is the remnant of a large extraterrestrial impact, the 40 km diameter Carswell Crater. When the meteor struck at this location, at around 115 Ma, the impact and subsequent rebound of the crust was enough to bring metamorphic rock up to surface from beneath about 2,000 m of Athabasca Group sandstone. There is no connection between the Carswell Crater and the much older (~1.2 Ga) uranium deposits.

The Taltson Magmatic Zone (TMZ), which forms the boundary between the Churchill and Slave Cratons, consists primarily of granitic rock. One interpretation is that the TMZ formed along a convergent boundary, although this is not universally accepted.

The Slave Craton is dominated by granitic rocks and metamorphosed clastic sedimentary rocks. On its western edge, there is a large area of very old gneissic rock that includes the Acasta Gneiss, dated at 4.03 Ga, which, for the time being at least, is the oldest rock in the world (Figure 21.6).

Figure 21.6 A sample of the Acasta Gneiss on display at the Natural History Museum in Vienna [https://commons.wikimedia.org/wiki/File:Acasta_gneiss.jpg]

The Wopmay Orogen, interpreted as the site of another ancient continent-continent collision, lies to the west of the Slave Craton. Although mostly composed of felsic igneous rocks and gneisses, the Wopmay Orogen includes a body of mafic and ultramafic igneous rock called the Muskox Intrusion. Derived from a mantle plume and dated at about 1.1 Ga, the Muskox is comparable to a handful of other mafic and ultramafic intrusions around the world in that it has distinctive repetitive layering caused by settling of heavy metal-rich minerals within the low-viscosity magma. Muskox has high levels of nickel, copper, and chromium, and has the potential to have platinum and palladium like a similar body in South Africa. Ultramafic intrusions like Muskox do not take place on Earth today because the mantle is no longer hot enough.

The oldest rocks in British Columbia are the strongly metamorphosed sedimentary, volcanic, and intrusive rocks of the Monashee Complex, situated to the west of the Columbia River near Revelstoke (Figure 21.7). Aged around 2 Ga, these may actually be part of Laurentia.

Figure 21.7 Precambrian rocks in southern B.C. and Alberta [By SE after: http://geoscan.nrcan.gc.ca/starweb/geoscan/servlet.starweb?path=geoscan/fulle.web&search1=R=208175]

There are much more extensive Precambrian rocks within the Columbia and Rocky Mountains of southeastern B.C. and the southwestern corner of Alberta. The rocks of the Purcell Supergroup (a **supergroup** comprises more than one group) are present in the extreme southeastern corner of B.C. and adjacent Alberta, and extend well into the United States (as the Belt Supergroup). These are mostly unmetamorphosed clastic rocks deposited in rivers and lakes during the middle Proterozoic, at around 1,400 Ma, while Laurentia was still part of the supercontinent Columbia. When Columbia rifted apart, the division happened within the area of the Purcell/Belt rocks. Similar rocks of the same age are present in Tasmania and Siberia, and it is postulated that they were once part of the same depositional basin.

Exercises

Exercise 21.2 Purcell Rocks Down Under?

This map shows the geology of the Australian state of Tasmania. Identify which rocks might be comparable to the Purcell rocks of B.C. and Alberta.

In what way are these rocks different from those in Canada?

[from https://commons.wikimedia.org/wiki/File:Tasmania_simple_geology_map.png]

The Windermere Group rocks — also mostly clastic sedimentary — were deposited in the ocean along the western edge of Laurentia (Figure 21.7) in the late Proterozoic (around 700 Ma) after the breakup of Columbia. In fact, sedimentary rocks of this age extend all along the western side of the Rocky Mountains, well into Yukon. Deposition in this area was taking place during the late Proterozoic Snowball Earth glaciations, as can be seen in Windermere Group rocks of the Toby Formation from the area south of Cranbrook, B.C. (Figure 21.8). The Toby Formation is a fine-grained marine rock (mudstone) with numerous large angular clasts of limestone and quartz. The mud was deposited in the quiet water of a continental slope environment, and the large clasts were dropped from floating ice derived from glaciers on Laurentia. The Toby Formation is unique in this area; most of the rest of the late Proterozoic clastic sedimentary rocks in this region do not have glacial dropstones.

Figure 21.8 Late Proterozoic Toby Formation mudstone with glacial dropstones south of Cranbrook, B.C. [SE]

21.3 Western Canada during the Paleozoic

At the beginning of the Paleozoic (542 Ma), Laurentia was near the equator (Figure 21.2) and sedimentation was continuing on all of Laurentia's marine margins, including the passive margin (not tectonically active) on what is now the west coast. The clastic sediments of the Windermere Group are succeeded by mostly limestone beds (represented by the blue areas in Figure 21.7) interbedded in some areas with mudstone and sandstone. The most famous Cambrian rocks in the Rockies are those around Field, B.C., within Yoho and Kootenay National Parks. The Burgess Shale of the Stephen Formation is considered by some to be the most important fossil bed in the world because of its spectacular preservation of detail in a wide array of organisms that are ancestors to many of today's organisms and are not present in earlier rocks. The Walcott Quarry, on the pass between Mt. Field and Wapta Mountain has been known and studied for over 100 years (Figure 21.9). In 2012 a new Burgess Shale discovery was made at Marble Canyon, about 30 km to the southeast, by a team led by the Royal Ontario Museum (ROM). Fossils with similar levels of preservation are present, and several previously unknown organisms have been found. The ROM continues to work in the Marble Canyon area and some of their discoveries are described and illustrated on this website: https://www.rom.on.ca/en/blog/mighty-burgess-shale-fossil-site-discovered-in-kootenay-national-park. The Paleozoic strata of the Rockies also include Ordovician, Devonian, Carboniferous, and Permian sedimentary rocks. For example, Carboniferous limestone makes up most of the upper part of Crowsnest Mountain in southern Alberta (Figure 21.1).

Figure 21.9 The Cambrian Burgess Shale at the Walcott Quarry, Yoho Park, B.C., with Wapta Mountain in the upper right. [SE]

While clastic and carbonate sediments were accumulating along the western edge of Laurentia, much of the interior of the continent was submerged under inland seas that were connected to ocean most of the time. This region is known as the Western Canada Sedimentary Basin (WCSB). The Paleozoic sediments that accumulated within this basin show up as the blue areas in Figures 21.4 and 21.5; however, their extent is much wider than that because Paleozoic sedimentary rocks also underly the Mesozoic rocks within most of the areas that are light green on those maps. By way of example, a schematic cross-section through the Paleozoic and Mesozoic rocks of southern Manitoba is given in Figure 21.10. The section extends from the Saskatchewan-Manitoba border on the left to just east of Winnipeg on the right, and shows the Paleozoic rocks overlain on the rocks of the Precambrian Superior Craton.

Fifteen different Paleozoic formations, ranging in age from Ordovician to Carboniferous, are shown in Figure 21.10. Of these, 11 are dominated by carbonate rocks (limestone or dolomite) that very likely formed in an ocean-connected marine environment. The non-carbonate formations are the lowermost one (resting on Precambrian rocks), which is sandstone of marine origin; the Devonian Prairie Evaporite Formation (in red) — the same formation from which potash is mined in Saskatchewan; and the upper two Devonian formations (in yellow), which are shale. When the Prairie Evaporite formed, the basin was isolated from the open ocean, and the rate of evaporation was greater than the rate of input from precipitation and river inflow. During that time, probably at least several million years, there were numerous changes in sea level or land level that allowed additional ocean water — and therefore additional salt — into the basin.

Figure 21.10 The Paleozoic sedimentary rocks of southern Manitoba along a section extending from the Saskatchewan border on the left to the Winnipeg area on the right. The section is 400 km wide and 1,800 m high, and the vertical exaggeration is about 100 times. The dip of the beds is also exaggerated by 100 times; their original and current attitudes are close to horizontal. [SE]

There are Paleozoic rocks in the central and western parts of British Columbia and Yukon, but they formed far away and did not become part of North America until the Mesozoic. Subduction started along the western edge of Laurentia by the middle Paleozoic. That meant that oceanic crust was moving toward the continent, bringing small segments of exotic continental crust with it (Figure 21.11). These crustal blocks along western North America are called **terranes**, indicating that they are sections of the continent that have an exotic origin (Figure 21.12). Most of British Columbia is made up of terranes that include sedimentary rocks with fossils that imply an origin south of the equator, or volcanic rocks with magnetic orientations that indicate a southern-hemisphere origin.

Figure 21.11 The distribution of continents in the early Carboniferous, showing the terranes that later became attached to the west coast of North America. The light blue areas are continental shelves, the white is ice of the Karoo Glaciation, and the red line shows subduction of oceanic crust beneath Laurentia. Panthalassic is the name for the huge ocean that preceded the Pacific Ocean. [SE based on information from Christopher Scotese at http://www.scotese.com/]

Figure 21.12 The Carboniferous Mt. Mark Formation on Vancouver Island is part of the Wrangellia Terrane, which arrived on the edge of North America during the Cretaceous. [SE]

Exercises

Exercise 21.3 What Is Vancouver Island Made Of?

This map shows the main geological features of the Wrangellia Terrane rocks that were present when Vancouver Island arrived on the coast of North America.

1. Roughly what percentage of Vancouver Island is underlain by Paleozoic rock?

2. What is the most common age and type of rock on Vancouver Island?

To answer these questions, you might find it useful to fill in each rock type area using coloured pencils. [SE]

21.4 Western Canada during the Mesozoic

The Mesozoic extends over 187 million years from the beginning of the Triassic (252 Ma) to the end of the Cretaceous (65 Ma). It was a particularly important period for the geology of western Canada. During this time, several continental collisions occurred along the west coast, resulting in the formation of the Rocky Mountains and the **accretion** (addition) of much of the land mass of British Columbia, and continuing deposition within the WCSB.

Terrane Accretion in British Columbia and Yukon

Continued subduction along the western edge of North America carried a number of continental terranes toward the coast, with the first collisions taking place in the early part of the Triassic, as the Quesnel, Cache Creek, and Stikine Terranes combined to form the Intermontane Superterrane, so named because it forms the interior plateau of British Columbia, between the Rockies to the east and the Coast Range to the west (Figure 21.13).

Figure 21.13 Model of the accretion of the Intermontane and Insular Superterranes to the west coast of North America during the Mesozoic. Subduction zones are the red-toothed lines. The dark-red triangles represent volcanoes. [SE]

Approximately 100 million years later, another pair of terranes — Alexander and Wrangellia — collided to form most of Vancouver Island and Haida Gwaii, plus a significant part of Alaska. During the Cenozoic, additional terranes (the Outboard terranes) were added to the western edge of North America. An overview of the accreted terranes of B.C., Yukon, and Alaska is given in Figure 21.14.

Figure 21.14 A generalized overview of the accreted terranes of B.C., Yukon, and Alaska. The Intermontane terranes are in green, the Insular terranes in purple, and the Outboard terranes in yellow. The Coast Plutonic Complex (CPC) formed in situ and is not a terrane. [SE after Yukon and BC Geological Surveys]

During the Jurassic, the Intermontane Superterrane acted like a giant bulldozer, pushing, folding, and thrusting the existing Proterozoic and Paleozoic west coast sediments eastward and upward to form the Rocky Mountains (Figure 21.15). The same process continued into the Cretaceous as the Insular Superterrane collided with North America and pushed the Intermontane Superterrane farther east. Folding in Rocky Mountain rocks, like that shown in Figures 21.16 and 21.17, is one of the results of this process.

Figure 21.15 Cross-section of the accretion of the Intermontane Superterrane to the west coast of North America and the resulting compression, folding, and thrusting of North American sedimentary rocks. In the Late Cretaceous, it was the accretion of the Insular Superterrane pushing against the Intermontane Superterrane that did most of the work. [SE]

Figure 21.16 Tight folds in sedimentary rocks of the Rocky Mountains near Field, B.C. [SE]

Figure 21.17 Folding of sedimentary rocks at Mt. Rae, Alberta. [https://upload.wikimedia.org/wikipedia/commons/e/e5/Mt-Rae-Alberta-Canada-aerial1.jpg]

Thrusting is another important process in the formation of fold-belt mountains, as described in Chapter 12. During plate convergence, entire sheets of sedimentary rock are slowly pushed over top of other sheets, resulting in situations where older rocks lie on top of younger ones. One of the best known examples of this is at Mt. Yamnuska, near Exshaw, Alberta (Figure 21.18), where the older Cambrian rocks were pushed east by a total of 40 km, over top of younger rocks.

Figure 21.18 The McConnell Thrust at Mt. Yamnuska near Exshaw, Alberta. Carbonate rocks of Cambrian age have been thrust over top of Cretaceous mudstone. [SE]

In the area near the U.S. border, within B.C., Alberta, and Montana, a sheet of Paleozoic rocks has been thrust about 80 km east over top of Cretaceous rocks along the Lewis Fault (Figure 21.19).

Figure 21.19 The Lewis Thrust at Crowsnest Mountain near Frank, Alberta. Carbonate rocks of Devonian and Carboniferous age have been pushed 80 km to the east and thrust over top of Cretaceous mudstone. [SE]

Not only did the subduction of oceanic crust beneath North America during the Mesozoic deliver geologically exotic terranes to the western edge of the continent, it also resulted in massive amounts of volcanism along the boundary (Figure 21.13). The upper-crustal magma chambers that fed those now-eroded volcanoes slowly cooled into granitic and dioritic stocks, and those stocks gradually coalesced into batholiths that extend from the southwest corner of B.C. all the way into Yukon and Alaska (Figures 21.20 and 21.21). Most of the granitic rocks of this region fall into two main age ranges: many are middle-Jurassic to early Cretaceous in age (~ 170 Ma to 140 Ma), while others are late Cretaceous to Paleogene (~50 Ma to 90 Ma). Many of the older bodies intruded into the terranes they are on before they arrived on the North America coast. This applies to those on Vancouver Island and Haida Gwaii. Some of the older ones formed in situ when the subduction zone was farther east (see Figure 21.13). Most of the younger bodies formed in situ when the subduction zone was close to where it is now (west of Vancouver Island) or slightly to the east.

Although these intrusive igneous rocks cooled at depth in the upper crust, they now form some of the highest peaks in Canada, many of them hundreds of metres higher than those of the Rocky Mountains (including Mt. Waddington, the highest peak entirely within British Columbia at 4,091 m). It is estimated that over the past 100 million years some of these igneous bodies have been uplifted in the order of 8,000 m. Much of that uplift is a result of the relative low density of the granitic rocks compared with the surrounding rocks.

Figure 21.20 Jurassic (J) and Cretaceous (K) granitic bodies in south-central B.C. [SE]

Figure 21.21 Cretaceous granite at The Lions, north of Vancouver, with Howe Sound in the background [Isaac Earle, used with permission]

The Western Canada Sedimentary Basin during the Mesozoic

The construction of the Rocky Mountains during the Jurassic and Cretaceous — and their ensuing erosion — created a significant new source of sediments for the WCSB. Based on the ages, distributions, and thicknesses of the sedimentary layers (Figure 21.23), it is evident that the greatest volumes of sediment were produced in the Upper Cretaceous (100 Ma to 65 Ma) and into the Paleocene (65 Ma to 55 Ma). The sediments accumulated in a basin that is thought to have been at least partly formed by the presence of a subducting slab of oceanic lithosphere underneath this part of North America. However, the ongoing uplift of the Rockies through this time period also led to isostatic depression of the crust. The western edge of the basin, which has about 4,500 m of Mesozoic rock alone, is a foreland basin. (See Chapter 6 for more on the origins of basins.)

Figure 21.22 The distribution of Mesozoic sedimentary rocks in the Western Canada Sedimentary Basin. J stands for Jurassic and Pc for Paleocene. The lines of the cross-sections of Figures 21.10 and 21.23 are shown. [SE after Alberta Geological Survey]

From time to time during the Mesozoic, the WCSB was filled to varying degrees with marine water. The Jurassic rocks at the base of the sequence are marine in origin, although the Jurassic sequence in Manitoba also includes evaporite layers. Most of the Middle Cretaceous rocks across the basin are marine, but the majority of the Upper Cretaceous rocks are of terrestrial origin, deposited within the flood plains and deltas of rivers. Some of these terrestrial sediments include coal layers; as described in Chapter 20, there are significant coal deposits in central Alberta.

The Paleozoic sediments within the WCSB were buried deeply beneath the Mesozoic sediments and were heated enough to form both oil and gas. There are large petroleum resources in reservoir rocks of various ages extending from northeastern B.C. to southwestern Manitoba.

Figure 21.23 A cross-section showing the Mesozoic sedimentary rocks in the WCSB from the Rocky Mountain foothills to north-central Saskatchewan [SE after Alberta Geological Survey]

Several of the terrestrial Cretaceous formations in the WCSB are host to important dinosaur fossils. Some are

within B.C. and Saskatchewan, but the most famous are in Alberta, including the Dinosaur Park Formation (Figure 6.1), the Scollard Formation, and the Horseshoe Canyon Formation (Figure 21.24). The Dinosaur Park Formation has one of the greatest concentrations of dinosaur fossils of any rock on Earth, with at least 50 genera of dinosaurs represented, ranging from tiny *Hesperonychus* to giant *Albertosaurus.* The Hilda Bone Bed, situated about 80 km to the east of Dinosaur Park, is estimated to have the remains of approximately 1500 ceratopsians, all of which are interpreted to have died in a flood related to a tropical storm. A few of the larger herbivorous dinosaurs found at Dinosaur Park are illustrated in Figure 21.25.

Figure 21.24 The Horseshoe Canyon Formation near Drumheller, Alberta [SE]

Figure 21.25 Depiction of some of the large herbivorous dinosaurs from the upper part of the Dinosaur Park Formation. Left to right: the ceratopsian Pentaceratops , the hadrosaur Lambeosaurus eating from a tall tree, the ceratopsian Styracosaurus, the ankylosuar Scolosaurus, the hadrosaur Prosaurolophus (in the distance), the ankylosaur Panoplosaurus, and a herd of Styracosaurs in the background. [by J.T. Csotonyi at https://upload.wikimedia.org/wikipedia/commons/e/e9/Dinosaur_park_formation_fauna.png]

Exercises

Exercise 21.4 Dinosaur Country?

The photo is from the Cretaceous Dinosaur Park Formation in Alberta. What characteristic of this rock tells you that it might be suitable for the discovery of the fossils of dinosaurs or other terrestrial organisms? [SE]

Several important depositional basins existed in British Columbia during the Mesozoic, including the large Jurassic-aged Bowser Basin north of Terrace, and the smaller late-Cretaceous Nanaimo Basin between Vancouver Island and the mainland. In both cases, the rocks are mostly clastic, with both terrestrial and marine deposition.

21.5 Western Canada during the Cenozoic

Two additional relatively small terranes collided with North America early in the Cenozoic. At around 55 Ma, metamorphosed sedimentary and volcanic rocks of the Pacific Rim Terrane were forced a few tens of kilometres underneath the west coast of Vancouver Island (Figure 21.26). These rocks are distributed along the west coast of the island and in the area around Victoria (Figure 21.27). At around 42 Ma, sea-floor pillow basalt and gabbro of the Crescent Terrane accreted to the southern margin of Vancouver Island and also to the adjacent part of Washington State. These terranes are shown as Outboard terranes in Figure 21.14.

Figure 21.26 East-west cross-section showing the accretion of the Pacific Rim and Crescent Terranes beneath Vancouver Island, and the ongoing subduction of the Juan de Fuca Plate. The dashed lines are inactive faults. [SE after Geological Survey of Canada]

Figure 21.27 The distribution of Pacific Rim and Crescent Terrane rocks on Vancouver Island [SE after Geological Survey of Canada]

The accretion of the Pacific Rim and Crescent Terranes had the effect of pushing Vancouver Island closer to the North American mainland, resulting in the uplift of the sediments deposited within the Nanaimo Basin to form islands in the Strait of Georgia (Figure 21.28) and mountains on Vancouver Island.

Figure 21.28 The Geoffrey Formation of the Nanaimo Group on Ruxton Island, B.C. [SE]

Following these events, the subduction of the Juan de Fuca Plate, which is a remnant of the former, much larger, Farallon Plate, was re-established at its current location farther to the west of Vancouver Island. This subduction, and that of the North America Plate beneath Alaska, has produced recently active volcanoes in Alaska, and all along the west coast from north of Vancouver Island to northern California (Figure 21.29). In southwestern B.C., there are several dormant volcanoes of Pleistocene age (including Garibaldi and Meager) that trend along a line that also passes through Mt. Baker in Washington State. About 40 km to the east is a trend of slightly older igneous complexes (Pliocene to Oligocene). The displacement between these belts could be explained by a westward shift in the position of the subduction zone over that time period.

Figure 21.29 The current plate situation along the western edge of northern North America. Blue lines are divergent boundaries, red lines are transform boundaries, and black lines with teeth are subduction boundaries. The dark red triangles are volcanoes. [SE]

The subduction and transform boundaries along this coast also generate relatively frequent earthquakes throughout this region, as illustrated in Exercise 11.1.

Figure 21.30 Oligocene to Pleistocene igneous complexes and volcanoes in southwestern B.C. and adjacent Washington [SE]

Sedimentation in the WCSB continued into the Cenozoic (Figure 21.22) with deposition of the Paskapoo Formation adjacent to the Rockies in Alberta (Figure 21.31), the Ravenscrag Formation in the Cypress Hills of southern Alberta and Saskatchewan, and the Turtle Hills Formation in southern Manitoba. All of these strata were deposited in terrestrial fluvial and deltaic environments, and all of them include coal deposits. Numerous mammalian (and other) fossils have been found in these rocks in Alberta and Saskatchewan. The mammals include primitive ungulates (ancestors to the deer and their relatives), a type of pangolin, a colugo (a gliding mammal that was possibly a primate ancestor), and some true primates in the suborder Plesiadapiformes, which became extinct and are not ancestors to any modern primates.

Figure 21.31 The Paskapoo Formation exposed on the banks of the Red Deer River, Alberta [https://commons.wikimedia.org/wiki/File:Paskapoo_Mudstones_Red_Deer.jpg]

Exercises

Exercise 21.5 The Volume of the Paskapoo Formation

The area underlain by the Paleocene Paskapoo Formation is outlined in yellow on the map shown here. The Paskapoo ranges up to about 1,000 m thick, and contours of its thickness (known as isopachs) are shown. The average thickness is about 500 m, and the area covered by the formation is about 90,000 km^2. This means that the rock has a volume of about 45,000 km^3. The sediments of the Paskapoo were derived from the Paleocene Rocky Mountains, within the area shown in blue, which is about 60,000 km^2.

1. What average depth of erosion would have been required within the source area to produce the 45,000 km3 of sediment, assuming that all of the eroded sediment ended up in the Paskapoo Formation?

2. The Paskapoo was deposited over a period of 4 million years from 62.5 Ma to 58.5 Ma. Assuming an average thickness of 500 m, what was the average rate of deposition (in mm/year) over that period?

[SE after Alberta Geological Survey]

Rocks younger than Paleocene (i.e., younger than 55 Ma) are relatively rare across the prairies, but there are widespread Eocene-aged volcanic and sedimentary rocks in central and southern B.C. The Kamloops Group includes the Tranquille Formation of lacustrine sediments (lake deposited), overlain by the Dewdrop Flats Formation of basaltic and andesitic volcanic flows and breccias. The Tranquille Formation includes the McAbee Beds and a number of other important sites with Eocene fossils (Figure 21.32).

Figure 21.32 The Tranquille Formation at the McAbee fossil site west of Kamloops, B.C. [SE]

The earliest Pleistocene glaciation in Canada started at about 2.64 Ma (late Pliocene) in the Klondike area of Yukon. This was part of the Cordilleran Ice Sheet. The Laurentide Ice Sheet started to form shortly afterward, and within 200,000 years had covered a large part of Canada and extended well into the United States. The Pleistocene glaciations had a major impact on the topography and geology of western Canada, creating extraordinary glacial erosion features in the mountainous regions of the west (Figure 21.33), and leaving enormous volumes of glacial sediment and glacial depositional features throughout the region (Figure 21.34).

Figure 21.33 Various glacial erosion features (top left) and glaciation at the Overlord Glacier, Coast Mountains, B.C. [Isaac Earle, used with permission]

Figure 21.34 A drumlin field with an esker (centre) in the Cree Lake area of northern Saskatchewan [NASA Landsat image, from Google Earth]

Chapter 21 Summary

The main topics of this chapter can be summarized as follows:

21.1	Geological History of Canada	The continent Laurentia, which includes what is now the Canadian Shield, was formed through the assembly of a number of smaller continents over the period from around 4 Ga to 1 Ga. Over the past 650 Ma, Laurentia moved north from deep within the southern hemisphere. During that time, a number of important plate-tectonic events have taken place, including formation of the Appalachian and Innuitian fold belts, sedimentation within the interior of the continent, and terrane accretion and mountain formation along the west coast.
21.2	Western Canada during the Precambrian	The two oldest parts of the Canadian Shield in western Canada are the Slave and Superior Cratons, and both include some of Earth's oldest rocks. During the formation of Laurentia, these cratons were combined with the Rae and Hearn Cratons, with the collision zones now represented by the Trans-Hudson Orogen and the Taltson Magmatic Zone. Continental sediments accumulated in Saskatchewan and Nunavut at around 1,700 Ma and in the area of the southern B.C.-Alberta border at around 1,400 Ma. This region was rifted apart during the breakup of the supercontinent Columbia, after which sedimentation continued on the western margin of Laurentia with the deposition of the Windermere Group.
21.3	Western Canada during the Paleozoic	Sedimentation on the west coast of North America continued into the early Paleozoic, and by the Ordovician, the Western Canada Sedimentary Basin (WCSB) had developed, extending from southern Manitoba to the northern Northwest Territories. Most of the Paleozoic rocks in this basin, which range in age up to the Carboniferous, have marine affinities, although there are important evaporites as well. Sedimentation continued on the west coast through this time, but by the late Paleozoic, a subduction boundary had developed along the coast and small continents were moving toward North America.
21.4	Western Canada during the Mesozoic	The various parts of the Intermontane Superterrane began colliding with the west coast of North America during the Jurassic (~180 Ma). This started the building of the Rocky Mountains by the thrusting of existing sedimentary rocks toward the east. The arrival of the Insular Superterrane during the Cretaceous (~90 Ma) contributed to further thrusting and uplift of the Rockies, creating a significant source of sediments for the WCSB. The greatest volume of Mesozoic rocks in the basin are of Upper Cretaceous age, and that likely coincides with the period of maximum collision-related uplift of the Rockies. Subduction of oceanic crust at various locations along the west coast and within the accreted terranes prior to their arrival has produced massive volumes of intrusive igneous rocks within the Coast Range.
21.5	Western Canada during the Cenozoic	The Pacific Rim and Crescent Terranes were added to the western edge of Vancouver Island during the Paleogene, pushing the island closer to the mainland and forcing recently deposited Nanaimo Group rocks onto the island. Continuing subduction along the coast has generated ongoing volcanism and earthquake activity in southwestern B.C. Sedimentation continued in the WCSB into the Cenozoic, especially in the Paleocene with deposition of the terrestrial Paskapoo Formation in Alberta and similar rocks in southern Saskatchewan.

Questions for Review

1. What are the oldest parts of Laurentia?2. The five main geological regions of Canada are shown on this map. Name the regions A through E.

3. Which ancient continent collided with North America to form the Innuitian fold belt, and when did that take place?

4. Explain why the ancient sedimentary rocks of the Athabasca and Thelon Basins are generally unmetamorphosed and undeformed.

5. Explain why ultramafic intrusions, like those of the Muskox Intrusion, are relatively common in Archean rocks, but rare in Phanerozoic rocks.

6. Use the Internet to find out why Cambrian marine organisms are so well preserved in the rocks of the Burgess Shale of British Columbia.

7. The Prairie Evaporite Formation overlies marine carbonate rocks of the Winnipegosis Formation and is overlain in turn by marine carbonate rocks of the Dawson Bay Formation. What type of changes might have led to the accumulation of evaporites during this period of marine deposition?

8. What features of the Intermontane Superterrane have been used to indicate that these rocks formed south of the equator?

9. What is the connection between terrane accretion on the west coast and the relatively rapid accumulation of sediments within the WCSB?

10.Why is the WCSB considered to be a foreland basin during the Mesozoic?

11. The four main terranes of the Intermontane Superterrane are Cache Creek, Quesnel, Stikine, and Yukon-Tanana. Referring to Figure 21.14, determine the order in which these terranes are likely to have reached North America.

12. The presence of Nanaimo Group sedimentary rocks far inland and at relatively high elevations on Vancouver Island is attributed to the accretion of the Pacific Rim and Crescent Terranes. What is the likely connection?

13. Referring to the diagram in Exercise 21.5, explain why the Paskapoo Formation gets thinner toward the northeast.