Fundamentals of Earth Science I
Notes
2014-2015
Marc Humblet
Nagoya University
G30 Program
1. Introduction

1.1. The scientific method

Science is about understanding how nature works. It is carried out by making hypotheses based on observations and/or experiments. Hypotheses are confirmed, revised or refuted as scientists keep exploring nature, conducting new observations and performing new experiments. A set of hypotheses explaining some aspect of nature is called a theory.

**Darwin’s theory of evolution by natural selection**

One fundamental scientific achievement that greatly improved our understanding of nature is Charles Darwin’s (1809-1882) theory of evolution by natural selection*. Darwin’s theory explains how new species can arise and how life evolved from simple microbial forms more than 3.5 billion years ago to the complex animals and plants we see today. Darwin’s discovery is often referred to as the “Darwinian Revolution”. Stephen Jay Gould, a famous American paleontologist, summarizes Darwin’s theory in three points: (1) organisms belonging to the same species display variations (color, size…) inherited by their offspring, (2) organisms produce more offspring than can survive, and (3) as a consequence of (1) and (2), organisms with more advantageous traits have a greater chance to survive and reproduce (they are naturally selected).

Darwin tells us that if a population of organisms of a particular species becomes reproductively isolated from other members of the same species, the combination of new environmental constraints and random genetic mutations may lead to the formation of a new species after many generations. Darwin’s understanding of evolution came in part from his observations of species geographic distribution during his journey around the world on board HMS Beagle from 1831 to 1836. For instance, Darwin noticed a resemblance between the animals of the Galapagos Islands and those of the American continent. He hypothesized that the Galapagos Islands must have been colonized by organisms from America and that these organisms subsequently evolved in isolation to form new species by natural selection**. Darwin’s theory is also based on his study of domesticated pigeon breeds. He experimented with the artificial selection of different varieties of pigeons by crossing individuals with particular traits. He surmised that given enough time (after many generations) nature could produce different species much like a professional breeder can produce different varieties of domesticated animals, or farmers different varieties of fruits and vegetables.

The issue of time was crucial to Darwin’s theory. There was no consensus at that time about the age of the Earth and many still believed the Earth was very young, perhaps as young as several thousands of years, much too young for Darwin’s idea of natural selection. Darwin’s view of the geological time scale was influenced by a book he was reading on board HMS Beagle called “Principles of Geology” written by a prominent British geologist named Charles Lyell. This book recognizes the need of huge time

---

*Note that the same idea was proposed independently by another British naturalist: Alfred Russel Wallace (1823-1913).
** Darwin’s finches famously illustrate how birds originating from the same species can evolve into distinct species by adapting to various environments (e.g. species with pointed beaks for picking fruits, species with shorter, larger beaks for eating seeds on the ground).
spans to explain the slow and progressive changes shaping Earth’s landscape. This naturally appealed to Charles Darwin whose theory of evolution by natural selection requires long stretches of time (millions of years) to evolve new species by the slow and progressive accumulation of tiny modifications.

There are different ways of doing science. For instance, theoretical physicists do not draw conclusions from a set of data in the same manner that Darwin did to elaborate his theory. Theoretical physicists use the language of mathematics to explain how nature works. Their findings are then confirmed or refuted by observations and/or experiments. General, basic principles explaining how nature works are referred to as physical laws. Examples of such laws are Newton’s three laws of motion and his law of universal gravitation which explain the relationship between the motion of an object and the forces acting upon it.

When a natural system becomes relatively well understood, it is possible to create a scientific model explaining its behavior and predicting its response in different conditions. In geology for example, materials with different densities can be used to simulate rock layers in the lab and examine how they behave when exposed to different kinds of stress (analog modeling). Our understanding of natural systems can also be translated into computer models. Computer modeling has become a major tool to simulate natural processes. Computer models which are the most familiar to us are probably those used to predict the weather. We see the output of these models everyday in the news.

1.2. What is Earth Science?

Earth science is a multidisciplinary science and its purpose is to study the Earth. Geology is the study of the solid earth (e.g. its history, composition, internal structure, and surface features). Geology is composed of various disciplines. For example, paleontology is the study of past life based on fossils. Geophysics and geochemistry aims at understanding geological processes using tools and principles of physics and chemistry, respectively. Geology is based on the study of the geological record, i.e. the information preserved in rocks formed at various times. Data of geological interest are primarily collected through field observations, rock sampling, lab analyses, geological and topographic mapping, seismic surveys, aerial photography, and satellite remote sensing (e.g. gravimetry).

Other branches of Earth science include oceanography — the study of the oceans —, meteorology — the study of the atmosphere —, glaciology — the study of ice and ice-related processes —, and geobiology — the study of the interactions between the biosphere, the lithosphere and the atmosphere —. Note that all these disciplines are themselves multidisciplinary. Over the past few decades, progress in space exploration has enabled us to look at the planets of our solar system at an unprecedented level of details. Our understanding of the geological
processes occurring on Earth combined with a wealth of data collected by satellites and land-based missions (e.g. Mars exploration rovers) are key to the study of the planets of our solar system (planetary science).

1.3. A brief history of geology

More than 2000 years ago, Greek philosophers were already trying to solve geological problems. For instance, they had noticed the presence of seashells in sediment layers well above sea level. Erastosthenes (3rd century BC) suggested that sea level may have fallen after the opening of the Strait of Gibraltar. Strabo (ca. 63 BC – ca. AD 24), on the other hand, proposed that catastrophic events such as earthquakes, volcanic eruptions, or landslides, had resulted in repeated sea level changes.

Another question exciting the curiosity of Greek philosophers was related to the more fundamental issue of the evolution of the Earth. By observing nature, they knew that rivers can carve deep valleys. They surmised that the process of erosion of hard rocks by water must be slow and progressive and must act over an extensive period of time to incise large mountains. Hence, Greek philosophers had already a sense of the long duration of the geological time scale. Moreover, they pondered about the continuous destruction of rocks by erosional processes which implies that mountains are meant to disappear on the long run unless they are regenerated by some other processes. This idea relates to the cyclic nature of geological processes which is central to modern geology (see plate tectonics and the rock cycle explained in the following chapters). Different scenarios were imagined by different schools of philosophers to explain the evolution of the Earth (see related slide). Although these ideas were inspired by observations, Greek philosophers were not scientists because they did not try to test their hypotheses.

Centuries later, a French philosopher named René Descartes (1596-1650), best known for his mathematical legacy (Cartesian coordinates) and philosophical work (e.g. Discourse on the Method and his famous “cogito ergo sum” — I think therefore I am —), elaborated a theory on the formation of the Earth and the origin of mountains and oceans. Based on his theory, the Universe is composed of three types of matter. One is the matter that composes the stars. Another is the matter that composes terrestrial bodies like the Earth. The third is the matter that composes the sky. Stars evolve into planets by forming an external crust. The Earth is thus a former star and possesses a central fire, remnant of its stellar origin. Various layers surround the central fire, including an internal ocean. The Earth’s landscape — mountains and ocean basins — is the result of the irregular collapse of the solid external layer. Although this idea has been completely refuted by later studies, it influenced generations of scientists who favored the idea of collapse rather than uplift to explain the formation of mountains. We know now that mountain chains can form by uplift where two tectonic plates converge but the theory of tectonic plates and our modern understanding
of mountain formation came to light only 300 years after the death of René Descartes!
An important step in the history of geology is marked by the achievements of a Danish anatomist named Nicolas Steno (1638-1686) who is considered to be the first “true” geologist. In the 17th century, the biological origin of fossils was not yet accepted. Steno identified a previously mysterious fossil called Glossopetrae as being the remains of the teeth of once-living sharks. He was also the first to formulate two fundamental principles of geology: the principle of horizontality and the principle of superposition. The first states that sediments are originally deposited horizontally; the second that older layers are at the bottom and younger layers at the top. Steno also realized that different assemblages of fossils represent different environments, and therefore that fossils can be used to reconstruct past environments. This is a fundamental concept underpinning modern stratigraphy, the study of layered rocks.

A hot debate which kept geologists busy in the 18th and 19th century opposed the neptunists and plutonists. The first school of thoughts was led by a German geologist named Abraham Gottlob Werner (1749-1817) and the second by a British geologist named James Hutton (1726-1797). Based on the neptunistic viewpoint, the Earth was originally covered by a global ocean. Most rocks found on Earth formed under water, including granites and basalts. We know now that granites and basalts form instead by crystallization of a cooling magma. The neptunists believed that mountains were formed by a combination of erosional and depositional processes in the ocean. For them, Earth’s internal heat had little influence on the evolution of the landscape. Conversely to this view, the plutonists considered Earth’s internal heat as the main driving force behind the formation of mountains and ocean basins. They correctly recognized the link between granites, basalts, and volcanic activity.

Another popular theory of the 19th century is called contractionism. It is based on the assumption that the Earth’s interior is progressively cooling; as the Earth cools and contracts, the rigid external crust deforms. This deformation results in the creation of mountains and ocean basins, much like the wrinkles forming at the surface of a drying apple.

Another question that was central to geology in the 19th century was to know whether we can use our knowledge of present-day geological processes to understand the past. More simply put: is the present the key to the past? Two very different schools of thoughts emerged: uniformitarianism and catastrophism. The uniformitarianists, led by Charles Lyell (1797-1875), held the view that the history of our planet is characterized by geological processes that are still acting today. Hence, it is possible to understand the past based on a careful study of the processes shaping our modern landscape. The catastrophists, led by the French naturalist Georges Cuvier (1769-1832), emphasized the importance of catastrophic events which have no analogue with present-day phenomena to explain the evolution of our planet (cataclysmic earthquakes, volcanic eruptions or floods). Our modern understanding tells us that the basic idea behind uniformitarianism is true. We can indeed understand a great deal about the past based on our
knowledge of geological processes occurring today. For instance, we can recognize beach deposits or river beds in the fossil record by analogy with modern sediments. We can recognize old mountain chains (e.g. Appalachian Mountains) and understand how they formed based on the study of active plate boundaries (e.g. Andes, Himalayas). However, the catastrophists were not entirely wrong because we know now that exceptional events like the impact of a very large meteorite or huge volcanic eruptions never witnessed in historical times have happened in the distant past and will happen in the future. For instance, the most widely accepted and best documented cause of the disappearance of the dinosaurs 65 million years ago is the impact of a very large meteorite off the Yucatan Peninsula.

The Newtonian revolution in physics took place in the 17th century. Geologists also have their revolution but it happened only recently in the mid-20th century. Although various theories explaining the formation of mountains and ocean basins, earthquakes and volcanoes have been proposed since the Greek antiquity, it is only in the 1950s that we began to understand the mechanisms behind these phenomena. The story begins with the discovery of continental drift by Alfred Wegener (1880-1930). Wegener compiled evidence supporting the fact that the position of continents is not fixed. However, he had no clue regarding the mechanism driving this movement. The time of Wegener’s discovery is also the time at which radioactivity was discovered by Henri Becquerel (1852-1902). Soon after, Ernest Rutherford (1871-1937) came up with a technique to date rocks based on the decay of radioactive isotopes which confirmed the old age of the Earth. The link between Earth’s internal heat and radioactivity was soon established, which led a British geologist named Arthur Holmes (1890-1965) to propose that heat produced by radioactivity could generate convection movements beneath the Earth’s crust and drive continental drift. The final pieces of the puzzle were put together in the mid-20th century when bathymetric surveys of the ocean floor led to the discovery of mid-ocean ridges. Dating of the ocean crust on both sides of these ridges showed that the age increases with the distance from the ridge. Based on this and other supporting evidence, geologists formed a theory known as the theory of plate tectonics (for details about plate tectonics, see chapter 3 of this course). Based on this theory, the Earth’s external rigid layer is divided into plates. Plates are moving relative to one another. New ocean plate is formed at mid-ocean ridges where two plates are pulled apart. Mountain chains form where two plates converge. Volcanoes and earthquakes are primarily located along plate boundaries where tectonic forces are concentrated. The theory of plate tectonics is at the core of modern geology and has revolutionized our understanding of the Earth and its history.

References
2. Formation of the solar system

2.1. Outline of our solar system

There are 8 planets in our solar system (listed by increasing distance from the Sun): Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune. These planets can be divided into the inner planets or terrestrial (also rocky) planets (Mercury, Venus, Earth, and Mars) and the outer planets, also called the giant planets (Jupiter, Saturn, Uranus, and Neptune). Jupiter and Saturn are the gas giants whereas Uranus and Neptune are often referred to as the ice giants. The inner and outer planets are separated by the Asteroid Belt: a zone comprised between the orbits of Mars and Jupiter and composed of a large number of asteroids. Asteroids are irregularly shaped bodies with diameters of several hundreds of km or less. Pluto, once considered a planet, is now officially identified as a dwarf planet. The term “dwarf planet” was ascribed to several small planetary bodies discovered recently in the Kuiper Belt, a zone of the solar system located beyond the orbit of Neptune. The definition of a dwarf planet is based on the following criteria:
- It orbits the Sun.
- It is not a satellite.
- Its shape is rounded.
- Unlike “true” planets, its gravitational field is too weak to have cleared its orbit and objects of comparable size can share the same orbital path.

The Kuiper Belt is also characterized by thousands of smaller objects composed in large part of frozen volatiles (ices) and is thought to be a major reservoir of comets. Another dwarf planet is Ceres, the largest object (950 km in diameter) and the only dwarf planet of the Asteroid Belt. Objects in orbit around planets or asteroids are called (natural) satellites or moons (note that our Moon is written with a capital “M” to distinguish it from other moons).

2.2. Birth of the solar system: the Nebular Hypothesis

There are several important observations that should be explained by any attempt to understand how our solar system formed:

1. The planets are all moving in the same direction (counterclockwise) around the Sun.
2. The Sun is also rotating counterclockwise.
3. The orbits of the planets are nearly circular and lie in the same narrow plane which coincides with the Sun’s equatorial plane (with the exception of Mercury whose orbit has an inclination of 7º with respect to Earth’s orbital plane, i.e. the ecliptic).
4. Satellites in orbit around planets are revolving in the same counterclockwise direction.
5. The Inner planets are smaller, denser and mostly rocky; the outer planets are larger and mostly made of gas or ice.
Among several other hypotheses, the most widely accepted nowadays is the *Nebular Hypothesis*, proposed for the first time by the German philosopher Immanuel Kant (1724-1804). It states that the solar system formed out of a rotating cloud of dust and gas (or *nebula*) contracting under its own gravity.

Nebulae are composed of gas, primarily hydrogen (H\(_2\)) and helium (He), and solid particles (dust), such as silicates, one of the most common family of minerals on Earth, and ice particles (e.g. CO, CO\(_2\), H\(_2\)O). The density of nebulae is not uniform. Some regions are denser than others. In denser regions gravity sometimes begins to pull surrounding particles and the region in question contracts. As more matter is pulled toward the center of that zone, density increases and temperature rises. There are two important consequences of the contraction of a nebula:

1. **As the cloud contracts, it spins faster.** This is explained by the law of conservation of angular momentum (see slides for explanations). Ice skaters use the exact same law when they wrap their arms around their body to spin faster.

2. **As the cloud rotates faster, it flattens into a disk.** This is a consequence of the centrifugal force (inertia) which prevents the equatorial region of the rotating cloud from collapsing. The linear speed of particles in the equatorial plane being maximum and decreasing steadily toward the poles, the centrifugal force in this plane is also maximum and opposes the gravitational inward pull. At the poles however the linear speed is zero and particles fall freely toward the center of the cloud under the influence of gravity. To understand the effect of inertia on a spinning object, think about a spinning ball of pizza dough!

### 2.2.1. Formation of the Sun

The contracting nebula evolves into a fast-rotating, flattened disk, the center of which is occupied by a dense and bright star in formation. The temperature inside the proto-star is so high that elements exist in the form of atoms stripped of their electrons (plasma). When the temperature at the center of the disk reaches 12,000,000 °C, the hydrogen nuclei (protons) overcome electrostatic repulsion and begin to fuse. This is the beginning of *nuclear fusion*, a chain reaction producing a great amount of energy and responsible for our Sun’s heat. The start of nuclear fusion marks the birth of a new star.

### 2.2.2. Formation of the planets

The region of the disk surrounding the proto-star is rich in dust and gas. Dust particles collide and begin to aggregate into larger chunks of matter. Larger chunks attract each other by gravity and form still larger bodies called *planetesimals* (>1 km in diameter). These planetesimals attract each other and smaller objects crossing their path, ultimately forming the planets of our solar system. Since lighter, more volatile elements are blown away by the solar wind (stream of charged particles) and the heat produced by the proto-star, the planets forming in the outer region of the solar system are enriched in gas and ice. They are the outer planets, also referred to as the giant
planets (Jupiter, Saturn, Uranus, and Neptune). Conversely, the planets forming in the inner region of the solar system are composed mostly of rock consisting of refractory (resistant to heat) components (e.g. silicates, iron, and nickel). These planets are the inner planets, also called terrestrial or rocky planets (Mercury, Venus, Earth, and Mars). Asteroids are the left-over of planetary formation, chunks of matter which have not been incorporated into the planets or their satellites. Rocky asteroids are concentrated in the Asteroid Belt whereas most icy objects are confined to the outskirt of the solar system, i.e. the Kuyper Belt. Astronomers think that there may be another more distant belt of icy objects called the Oort Cloud which could also be a potential source of comets.

Since asteroids and comets are composed of the original raw material of the solar system which has not been modified by the process of planetary formation, they represent a tremendous source of information concerning the initial conditions of the solar system at the time of its birth. That is why so much effort has been put in the exploration of these objects (see slides). Conveniently for scientists, asteroids sometimes cross the Earth’s path and the fragments fallen on Earth’s surface (meteorites) can be collected for analysis. Since these meteorites are pristine material originating from the dawn of the solar system, they are used to date the age of the solar system and the planets. On Earth, plate tectonics has destroyed most of the oldest rocks which is why we cannot determine precisely the age of Earth based on terrestrial rock samples (see next chapter for an explanation of the theory of plate tectonics). The current accepted age of the Earth (and the solar system) is 4.56 billion years based on meteorites (see chapter "the age of rocks" for more information about the radiometric dating technique).

2.3. Formation of Earth’s layers and Earth’s atmosphere

2.3.1. Earth’s internal structure

The Earth is thought to have been in a molten state at the beginning of its history. The primary sources of heat were:
1. The impacts of meteorites and large planetesimals (conversion of kinetic energy)
2. Radioactive decay (still a source of heat today)
3. Gravitational contraction (think about the "bicycle pump effect")

The molten state of the Earth meant that matter could move around and redistribute according to its density. Heavier components sank toward the center of Earth whereas lighter components remained closer to the surface. This process is called gravitational differentiation and is responsible for the formation of Earth’s layers:

- The dense core has a diameter of 3,500 km and is mainly composed of iron and nickel. It is divided into an inner core (solid) and an outer core (liquid).
- The mantle is 2,850 km thick and its density is intermediate relative to the core and the crust. It is composed of silicates very rich in Fe and Mg. This layer is characterized by convection
movements (hot matter rising, cold matter sinking) which drive plate tectonics (see next chapter).

- The outermost layer is the **crust** which is thinner beneath the oceans (7 km) and thicker beneath continents (40 km). The former is called the **oceanic crust** and is denser, (silicates rich in Fe and Mg). The latter is called the **continental crust** and is lighter (quartz — SiO$_2$ — and silicates rich in Al, K and Na).

2.3.2. Earth’s atmosphere

One can think about two possible sources for the gases (including water vapor) which accumulated around the earth to form the atmosphere: (1) volcanic eruptions releasing gasses trapped in the Earth’s interior and (2) extraterrestrial input of volatiles from objects impacting the Earth (especially comets).

Most importantly, the primitive atmosphere did not contain free oxygen (O$_2$). Its composition was different from today. Oxygen was produced later by *photosynthesis* (see chapter on the origin and evolution of life). Moreover, temperature was initially too high for water to be present at the liquid state. Only when surface temperature dropped below 100 °C did water vapor present in the atmosphere condense in a liquid state. We know that between 3.7 and 3.8 billion years ago, a permanent ocean was present at the surface of Earth. Liquid water could probably be found earlier but the frequency of large meteorite impacts may have vaporized part or all of it repeatedly (see section 2.5.1).

2.4. The Moon

2.4.1. The formation of the Moon

Several theories exist regarding the formation of the Moon but the most widely accepted explanation is the **Giant Impact Hypothesis**. Several observations must be explained when one wants to understand how the Moon formed:

1. Earth is the only inner planet of the solar system with a large moon.
2. The orbital plane of the Moon does not coincide with the orbital plane of the Earth around the Sun or with the Earth’s equatorial plane.
3. The Moon has a remarkably small iron core compared to the inner planets.
4. The Moon is much less dense (3.3 g/cm$^3$) than the Earth (5.5 g/cm$^3$). The Moon’s density is close to that of the Earth’s mantle.
5. There are very little light, volatile elements (including water) on the Moon.

These observations indicate that the Moon may not have formed out of the same raw material which produced the Earth. The giant impact hypothesis states that a large planetesimal collided with Earth 4.56 billion years ago and that the Moon formed out of the material ejected during the impact. The impact would have vaporized all the lighter, more volatile elements, explaining their
rarity on the Moon. The material ejected during the impact may have come primarily from Earth’s crust and mantle, explaining the low density of the Moon and the relatively small size of its iron-rich core. If the Moon is the product of an impact between Earth and another small planet, then one might expect to find the chemical signature of the impactor in the composition of Moon rocks. However, studies have shown that rocks from Earth and from the Moon have similar chemical compositions. If the Giant Impact Hypothesis is true, this implies that the small planet which collided with the Earth must have had a composition very similar to that of the Earth, perhaps because the impactor was formed in the same region of the nebula.

In addition to the formation of the Moon, the Giant Impact Hypothesis would also explain that the rotation axis of the Earth has a 23.5° tilt with respect to its orbital plane.

2.4.2. Characteristics of the Moon

The Moon is riddled with crater. Two types of surface can be identified:

1. Lowlands (Maria): surface with few craters consisting of basalt (rock formed by solidification of magma) that has filled larger craters.
2. Highlands: surface with many craters.

In planetary science, a very useful technique to determine the relative age of planetary surfaces is crater counting. In brief: the more crater, the older the surface. Hence, the surfaces called lowlands are younger than the surfaces called highlands. Scientists have been able to determine the exact age of these surfaces by dating the rocks collected on the Moon and brought back to Earth by the US Appolo missions and the Soviet Union Luna missions. The age of the highlands falls between 4.5 and 4 billion years (4.5-4 Ga) and that of the lowlands ranges from 4 to 3.2 billion years (4-3.2 Ga). The number of impact peaks at 4.5 Ga and again at 3.9 Ga (see graph on the corresponding slide). The first event is called the Heavy Bombardment, a period of very frequent meteorite impacts at the beginning of the formation of the solar system. The second event is called the Late Heavy Bombardment, another period of high-frequency meteorite impacts. Both events did not only affect the Moon but of course affected the Earth too!

2.5. Characteristics of the inner planets

The inner planets of the solar system must have originated from a similar raw material in the region surrounding the proto-Sun. Yet the inner planets differ greatly in many aspects: size, interior dynamics, atmospheric composition, and landscape (see slide with comparative table).

2.5.1. Mercury

Mercury is the planet closest to the Sun. It has almost no atmosphere and its surface is very old. The planet is therefore riddled by meteoritic craters like the Moon. Except for the occasional meteorite impacts, Mercury is geologically very quiet. Unlike the Earth there is no tectonic activity.
The surface is characterized by numerous escarpments which are thought to be faults formed shortly after the planet's formation when the crust contracted as it cooled. Surface temperature varies greatly depending on the exposure to sunlight.

2.5.2. Venus

Venus is characterized by an extremely dense atmosphere composed of 96% of CO\textsubscript{2}, the 4% left being primarily sulfuric acid and other gasses including water vapor. Due to its thick atmosphere, 80% of the incoming solar radiation is reflected back to space. However, the greenhouse effect due to the extremely high concentration of CO\textsubscript{2} is so intense that the surface temperature is maintained at a staggering 460 °C (enough to melt lead!). Venus is tectonically active and littered with volcanoes. The tectonics of Venus is probably very different from that of the Earth (described in next chapter). One hypothesis is that intense convection movements in the Planet’s interior generate many cracks at the surface breaking up the thin crust into flakes (hence the term “flake tectonics”). Unlike Earth, there are no large plates and no subduction zones. The surface of Venus is mostly composed of fractured plains and volcanoes. Due to volcanic activity, crust deformation, and shielding effect of the dense atmosphere, there are relatively few meteoritic craters.

2.5.3. Mars

Why is Mars red?
Mars is red because of the widespread presence of iron oxide (Fe\textsubscript{2}O\textsubscript{3}) and iron hydroxide (FeO(OH)) dust (rust) on its surface. These minerals form on Earth mainly by oxidation of iron-bearing silicates (e.g. (Fe, Mg)\textsubscript{2}SiO\textsubscript{4}). There is very little free oxygen (O\textsubscript{2}) on Mars and the reaction of oxidation is very slow. However, there is evidence indicating that free oxygen may have been more abundant in the distant past of the planet. But why is iron oxide much more common on the surface of Mars than on the surface of Earth? This problem has not been solved yet. One hypothesis suggests that the temperature of the Earth at the beginning of its formation was higher than that of Mars and that more iron oxide was melted and therefore more elemental iron migrated toward Earth’s core. On Mars the melting of iron oxide was not as complete due to the lower temperature and more iron oxide remained in the planet's outer layer.

Note: the reddish dust covering most of Mars surface is sometimes blown in the air by strong winds on a global scale during dramatic planet-wide dust storms!

Mars topography
From a topographic viewpoint, there is a striking difference between the northern and southern hemisphere. The northern hemisphere is characterized by lower elevations with few craters whereas the southern hemisphere is characterized by high elevations with many craters. One hypothesis involves a catastrophic impact with a large meteorite the size of a small planet which
would have completely altered the topography of the northern hemisphere.

Mars topography is also characterized by extremes in elevation. The tallest mountain is 25 km high (Olympus Mons, a large shield volcano). The deepest canyon is 8 km deep (Valles Marineris).

Water on Mars

The current conditions of temperature and atmospheric pressure on Mars are not favorable for the presence of liquid water on the planet’s surface. Water is present as vapor in the atmosphere but exists primarily as ice:

- in polar ice caps
- as subsurface ice (or ground ice) directly beneath the Martian soil
- as ephemeral frost depositing on the Martian soil in winter

Note 1: the Martian polar ice is composed of both water ice and “dry ice” (CO₂ ice). Dry ice sublimes in summer (seasonal ice) which causes large variations in atmospheric pressure.

Note 2: Satellite imagery has also revealed the presence of ancient glaciers at relatively low latitudes beneath a layer of Martian dust protecting them from sublimation. The presence of ice so far from the poles may indicate that the planet experienced ice ages characterized by larger ice caps than today.

There is undeniable evidence of liquid water flowing on the surface of Mars in the past. Three supporting lines of evidence exist:

1. Satellites orbiting Mars have produced stunning images of landforms clearly related to the presence of running liquid water such as meandering channels and fan deltas.
2. The rover Curiosity revealed the presence of layers of sedimentary rock composed of gravels analogue to river beds found on Earth. There are also images of sedimentary structures (layering, cross-stratifications) indicating the presence of sediments deposited under water.
3. Specific minerals thought to have formed in contact with liquid water have also been reported, such as clay minerals or the Martian spherules composed of hematite (Fe₃O₄).

Liquid water could occur on Mars today beneath the surface where temperature is higher due to the Planet’s internal heat. A “frozen sea” covered with what looks like rafts of sea ice lying beneath a layer of Martian dust was discovered a few years ago near the equator. Scientists think that the water composing this frozen sea came from a water-rich layer occurring underground and reached the surface a few million years ago for an unknown reason. The water quickly froze upon reaching the surface and was covered by dust. The dust layer prevents the ice from disappearing by sublimation. If life exists on Mars, reservoirs of groundwater beneath the planet’s surface may well be the place to find it!
3. The Earth system

Our planet is composed of three complex natural systems (Figs. 1 & 2): (1) the climate system, (2) the plate tectonic system, and (3) the geodynamo system. The first is driven by an external source of heat (i.e. the Sun), whereas the second and the third are driven by an internal source of heat (i.e. original heat accumulated during Earth’s formation and heat being generated by the decay of radioactive elements inside the Earth).

(1) The climate system involves all the components of the Earth whose interactions control Earth’s climate:

- **Atmosphere** (layer of gas around the Earth)
- **Cryosphere** (surface ice and snow, e.g. ice caps, glaciers)
- **Hydrosphere** (liquid surface waters, including groundwater)
- **Lithosphere** (rigid rocky outer layer of Earth)
- **Biosphere** (all living things on Earth)

Recent global warming calls to our attention the impact of human activities on climate. The anthropogenic influence on nature has become so great that it is nowadays adequate to define yet another component of the climate system: the anthroposphere (sum of all human activities influencing the environment).

(2) The plate tectonic system involves all the components of the Earth which control the movements of continents, the formation of mountains and ocean basins, and events such as volcanic eruptions and earthquakes. These components are the lithosphere, the asthenosphere, and the deep mantle. The principles of plate tectonics are described in section 3.2 (see slides for additional figures and further details on subjects related to plate tectonics).

(3) The geodynamo system produces and maintains Earth’s magnetic field and involves movements of charged particles in the liquid outer core.

---

**Figure 1**: Illustrations of the climate system (typhoon, A), the plate tectonic system (lava, B), and the geodynamo system (compass, C). Sources: A. NOAA, B. Encyclopaedia Britannica, C. Wikipedia.

**Figure 2**: Model of the Earth and the three systems: climate, plate tectonics, and the geodynamo.
3.1. The age and structure of the Earth

The exact age of the Earth cannot be determined directly but a good approximation can be obtained by measuring the age of meteorites. Meteorites are rocks falling from space on the surface of the Earth. They originate primarily from the collision and fragmentation of asteroids. Asteroids are the leftover of the process of planetary accretion. In other words they did not aggregate with other asteroids to form planets when the solar system formed. The age of the Solar System (and the Earth) derived from meteorites is 4.6 billion years.

During the first billion years, the solar system was not yet cleared of the majority of its asteroids which means there was a higher probability of collision among asteroids and planets, and therefore a higher number of meteorites falling on the Earth. This period of Earth’s history is referred to as the **Heavy Bombardment**. Heat energy released by the collision between Earth and large meteorites was high enough to melt Earth’s surface. Other important sources of heat were provided by gravitational contraction (think about the “bicycle pump effect”) and by the decay of radioactive elements contained in Earth’s interior. Consequently our planet was occasionally in a molten (“soft”) state during which elements could migrate freely, and matter redistributed according to its density. Heavier elements (Fe, Ni) migrated toward the center of the Earth to form the core, whereas lighter elements remained at the surface (Mg, Al, K). The process by which Earth became a layered planet is called **gravitational differentiation**.

Earth is composed of three main layers: the crust, the mantle and the core (Fig. 3A). Each layer has a distinct chemical composition. Temperature and pressure increase toward the center. Only the outer core is in a liquid state. The inner core is solid because the pressure is extremely high and “forces” matter into a solid state despite the very high temperature.

**Figure 3**: Schematic cross section of the Earth. (A) Earth’s main layers and (B) convection movements in the mantle (dark green = colder material sinking, yellow-orange = hotter material rising).

Most of the Earth’s volume consists of the solid mantle. Rocks of the crust and mantle are composed primarily of minerals of the silicate family. The basic structural unit of silicate minerals is \([\text{SiO}_4]^{4-}\) in which each oxygen atoms share one electron with the silicon atom. The core is Earth’s densest layer and composed mostly of an iron-nickel alloy. The mantle is slightly denser than the crust and composed of silicates very rich in Fe and Mg. The crust is divided into the oceanic and continental crust. The former is thinner (up to 7 km thick), heavier and enriched in Fe and Mg*. The oceanic crust lies beneath the ocean floor. The latter is thicker (up to 40 km thick), lighter and enriched in Al, K and Na (Fig. 4). The continental crust forms the continents and their continental shelves (shallow seabed adjacent to the coast). Where does the evidence supporting the model of

---

* Note that the chemical composition of the oceanic crust is close to that of the mantle because the oceanic crust is derived from the partial melting of mantle rock (see chapter on igneous rocks).
a layered Earth come from? There are 6 main sources of evidence: (1) the study of meteorites originating from larger bodies which have been subject to gravitational differentiation, (2) the study of seismic waves traveling inside the Earth, (3) the analysis of fragments of rocks from the upper mantle brought to Earth’s surface by magmas during volcanic eruptions, (4) deep drilling (project of drilling oceanic crust down to mantle by Japanese drilling vessel Chikyu), (5) high-pressure lab experiments, and (6) computer-based mathematical modeling.

Figure 4: Thickness, composition and density of Earth’s layers. Note the distinction between crust and mantle and between the lithosphere and the asthenosphere.

3.2. A dynamic Earth: the theory of plate tectonics

The crust and the uppermost part of the mantle form a rigid layer called the **lithosphere** (Figs. 4). The lithosphere beneath the oceans (i.e. **oceanic lithosphere**) has an average thickness of 70 km. The lithosphere that forms the continents (i.e. **continental lithosphere**) can be more than 200 km thick. Directly below the lithosphere lies the **asthenosphere** (or asthenospheric mantle), a layer of the mantle a few hundred kilometers thick. The physical properties of the lithosphere and asthenosphere are very different. The lithosphere is rigid and brittle. The asthenosphere is weak and ductile. Slow, plastic deformations can take place in the asthenospheric mantle, allowing matter to move around. The rock “flow” inside the mantle is driven by differences of density. Hotter material rises toward the surface whereas cooler material sinks (Fig. 3B). This vertical motion is called **convection** (like convection taking place in a hot miso soup in which denser, colder soup in contact with the bowl sinks and lighter, hotter soup rises until the temperature/density inside the soup becomes uniform).

The lithosphere is broken into plates called **tectonic plates** (Figs. 5 & 6). Tectonic plates move relative to one another and slide over the asthenosphere. The motion of plates is driven by convective movements inside the mantle. Three major types of plate boundaries can be identified: (1) divergent, (2) convergent, and (3) transform-fault.

Figure 5: The main tectonic plates. A, B and C are the 3 types of plate boundary: (A) divergent boundary, (B) convergent boundary, and (C) transform-fault boundary (note that only major transform faults are indicated by yellow arrows on this figure).

Figure 6: Global relief map of Earth. Note the correspondence between the location of plate boundaries and Earth’s topography. Source: NOAA.
(1) **Divergent boundaries**
Along divergent boundaries two plates are pulled apart and a large valley forms in between. This is a zone where mantle rock rises toward the Earth’s surface and partially melts*. Some of the molten rock (magma) solidifies before reaching the surface. Some reaches the surface and forms a volcanic chain in the middle of the valley. This process is responsible for the formation of new oceanic lithosphere along plate boundaries called **mid-ocean ridges** (e.g. the Mid-Atlantic Ridge, Figs. 6 & 7A). The divergence of oceanic plates at mid-ocean ridges and the production of new oceanic crust is a process referred to as **sea-floor spreading** (average spreading rate of mid-ocean ridges = 50 mm/yr). The other type of divergent boundary is the **continental rift** where continental crust is being pulled apart (e.g. the East African Rift, Fig. 7B) eventually leading to the formation of a new ocean basin.

(2) **Convergent boundaries**
Since the surface area of the Earth’s crust does not increase over time, the continuous production of oceanic lithosphere at mid-ocean ridges means that oceanic lithosphere must be destroyed somewhere else. This happens at **subduction zones** (Figs. 7C). At subduction zones, two plates converge (two oceanic plates or one oceanic and one continental plate). The heavier oceanic plate slips under the other one** and sinks deeper into the mantle where it is “recycled”. The subducting oceanic plate sinks because it is colder and denser than the surrounding asthenospheric mantle. Subduction zones are characterized by a deep **oceanic trench** on the subducting plate side (e.g. the Mariana Trench) and by a **mountain chain** on the overriding plate side (e.g. the Andes, Fig. 6). Subduction zones are also associated with volcanic activity*** (e.g. Mount Fuji). The other type of convergent boundary involves the collision between two continents: **continental collision** (Fig. 7D). In this case, the converging plates are both continental, hence light compared to the mantle, and no subduction can occur. Instead, a large mountain chain builds up where the two continents meet (e.g. the Himalayas, Fig. 6).

(3) **Transform faults**
The last type of plate boundary is called **transform-fault** (Figs. 7E-F). Along transform-fault boundaries, two plates slip past one another. They most commonly offset mid-ocean ridges but they can also be found on land (e.g. San Andreas Fault).

**Earthquakes** occur along all plate boundaries, whereas the majority of **volcanoes** are located along subduction zones and mid-ocean ridges.

---

* In this context, the partial melting of mantle rock is due to the decrease in pressure during its ascension toward the Earth’s surface (**decompression melting**). The melting temperature of a rock is lowered when pressure decreases.

** In case of a convergence between two oceanic plates, the older, cooler, hence denser plate subduces beneath the other one.

*** Rock melting at subduction zones is facilitated by the presence of H$_2$O contained in sedimentary rock pores and clay minerals. The effect of water is to lower the melting point of rocks because water helps break chemical bonds.
4. Minerals: rock’s elementary building block

4.1. What is a mineral?

A mineral is a naturally occurring solid inorganic substance composed of a regular 3D arrangement of atoms repeating in all directions (crystal structure). Each mineral can be identified with a chemical formula. A mineral growing freely form crystals whose external geometrical shape reflects the internal regular arrangement of atoms. The crystal structure of a mineral depends on its chemical composition. It also depends on the temperature and pressure at which it formed. For example, graphite and diamond have the same chemical composition (carbon) but they formed in very different conditions of pressure and temperature and therefore are characterized by very different crystal structures. Since diamond forms at very high pressure deep inside the Earth’s crust, the atoms of carbon in diamond are more densely packed than in graphite. Minerals with the same chemical composition but different crystal structures like diamond and graphite are called polymorphs.

Examples of minerals and their chemical formula:
Quartz: SiO$_2$, Olivine: (Fe, Mg)$_2$SiO$_4$, Halite: NaCl, Calcite: CaCO$_3$, Pyrite: FeS$_2$, Hematite : Fe$_2$O$_3$

4.2. Structure of matter and chemical bonds

Minerals are made of atoms. The simplest way to represent an atom is a nucleus composed of protons and neutrons surrounded by electrons occupying different orbits or shells. There is a specific number of electrons that can occupy each shells. Since atoms are electrically neutral, the number of protons in the nucleus is equal to the number of electrons. Each element is characterized by a specific number of proton(s)/electron(s) called the atomic number: H (1), C (6), N (7), O (8). Atoms of a given element characterized by different numbers of neutrons are different isotopes of that element. For example, the notation for an isotope of carbon that has 6 protons and 6 neutrons is carbon-12 or $^{12}$C. Two other isotopes of carbon are $^{13}$C (6 protons and 7 neutrons) and $^{14}$C (6 protons and 8 neutrons). There are two types of isotopes: stable isotopes and radioactive isotopes (or radioisotopes). The latter decay into stable or radioactive isotopes over time while emitting radiations. For example, $^{14}$C is radioactive and decays into the stable isotope $^{14}$N.

Since a neutron has a certain mass, different isotopes of a same element have a different mass. The mass of an atom can be approximated to the sum of masses of the protons and neutrons in the nucleus. By convention, the mass of the isotope of carbon $^{12}$C (6 protons and 6 neutrons) is 12 atomic mass units. The atomic mass of an isotope is the sum of the number of protons and neutrons in the nucleus. The atomic mass of an element depends on the relative abundances of its isotopes (e.g. the atomic mass of carbon is 12.011 because $^{12}$C is by far the most abundant isotope of carbon, i.e. 98.9% of all carbon).
The electrons participating in chemical bonds are usually those of the outermost shell. Chemical bonds allow the atoms to achieve a more stable electronic configuration. There is an optimal number of electrons on the outermost shell which can be achieved through chemical bonds. For example, the optimal configuration for C, N and O is to have 8 electrons occupying the outermost shell. In minerals, three types of chemical bonds exist: the *covalent bond*, the *ionic bond* and the *metallic bond*.

In **Covalent bonds**, an atom shares one or more electrons with another atom. The electrons of one atom are attracted by the positively charged nucleus of the other atom and vice versa. This configuration results in very strong chemical bonds.

*Example*: diamond in which each atom of carbon is surrounded by 4 atoms and each atom shares 4 electrons with its neighbors. Note that another mineral has the same chemical composition: graphite. However, graphite has a very different crystal structure. In graphite, the atoms of carbon are arranged in sheets in which atoms are held together by covalent bonds but the sheets are held together by a much weaker type of chemical bond (similar to the metallic bond described below). This makes graphite much less hard than diamond. The softness of graphite can be observed every time one uses a pencil to write or to draw. Pencil cores are made of graphite.

In **Ionic bonds**, an atom gains one or more electrons from another atom (transfer of electron(s)). An atom that has lost one or more electrons is called a *cation* (positively charged ion). An atom that has gained one or more electrons is called an *anion* (negatively charged ion). The ionic bond results from the attraction between ions of opposite charges.

*Example*: halite NaCl (table salt) in which each atom of chlorine “takes” one electron from an atom of sodium. The atom of chlorine that has gained one electron is negatively charged (anion) and attracted by the positively charged sodium (cation).

In **Metallic bonds**, atoms which have a strong tendency to lose electrons (cations) are surrounded by free-moving electrons (delocalized electrons). The cations are held together by their attraction to the surrounding electrons.

*Example*: copper which has one electron on the outermost shell that is only weakly attracted by the nucleus. In the metal copper, these outer electrons are not bounded to the nuclei and move freely. The cations Cu⁺ are held together by their attraction to the free electrons surrounding them.

The most common mineral family on Earth is the family of **silicate minerals**. Silicate minerals are often characterized by a combination of covalent and ionic bonds. The basic structural unit of silicate minerals is the silicon-oxygen tetrahedron or \([\text{SiO}_4]^4-\) (silicate anion) in which one atom of silicon is surrounded by 4 atoms of oxygen and shares one electron with each oxygen. Each atom of oxygen can share one electron with another atom of silicon (covalent bond) or bind with positively charged ions (ionic bonds).

Examples:

(1) Quartz (\(\text{SiO}_2\)) is one of the most common silicate minerals of the continental crust. Quartz is composed of a juxtaposition of silicon-oxygen tetrahedra. Each atom of oxygen is linked to 2
silicon atoms by electron sharing. There is only one type of chemical bond in this mineral: covalent bonds.

(2) Olivine is a silicate rich in Mg and/or Fe ((Fe, Mg)$_2$SiO$_4$) very abundant in the upper mantle and common in the oceanic crust. In olivine, the silicon-oxygen tetrahedra are linked to the metallic cations Fe$^{2+}$ or Mg$^{2+}$ by ionic bonds. The -4 charge of each silicate anions is balanced by the +2 charge of two metallic cations. There are two types of chemical bond in this mineral: covalent and ionic bonds.

There are many other combinations of silicate anions and other elements which produce a broad variety of crystal structures in the family of silicate minerals (e.g. single-chain silicates, double-chain silicates, sheet silicates…).

4.3. Crystallization processes

**Crystallization** is the process by which atoms in a gas or a liquid assemble in an orderly 3D pattern (crystal lattice) to form a solid substance (mineral crystal).

Crystallization occurs typically in the following three contexts:

(1) **Saturation**: When dissolved anions and cations in an aqueous solution reach a certain concentration, they become saturated and if water is removed by evaporation or if more ions are added to the solution the anions and cations combine to form crystals. One example is the precipitation of salt (NaCl = halite = table salt) when seawater evaporates in a pond, lake or embayment. The rock produced by accumulation of minerals precipitated from saturated brines is called *evaporite* (see chapter on sedimentary rocks).

(2) **Phase change** (liquid-solid, gas-solid): Crystals can form when the temperature of a liquid drops and the liquid solidifies into a solid — **solidification**. This happens when water freezes to form ice. It also happens when magma (molten rock) cools and minerals crystallize to form a solid rock, i.e. igneous rock (see chapter on igneous rocks). The transformation of a gas into a solid without transiting by the liquid phase is called *deposition*.

(3) **Biomineralization**: Many organisms are capable of precipitating minerals to form a skeleton or shell. Our bones and teeth are mineralized structures. In the marine world, the most common biomineral is calcium carbonate (CaCO$_3$). It is produced by diverse organisms such as reef corals, coralline algae, mollusks, urchins, sea stars, and foraminifera (one kind of large single-celled organism). Another common biomineral is silica (SiO$_2$) and is produced by several groups of marine microorganisms such as radiolarians and diatoms.

4.4. Common families (or classes) of minerals

The most common minerals are the *silicate minerals*. They form most of the Earth’s mantle and crust. I already provided some details about their structure.
Another class of minerals comprises the **carbonates**. The most common carbonate mineral is CaCO$_3$ already mentioned above in relation with biological mineralization. The basic structural unit of carbonates is the carbonate anion (CO$_3$)$^{2-}$. In carbonate minerals, carbonate anions are arranged in sheets and are bound to cations located between the sheets. In the case of CaCO$_3$, the cations are Ca$^{2+}$. Calcium carbonate can also precipitate abiologically in seawater saturated in Ca$^{2+}$ and (CO$_3$)$^{2-}$ or even (in relatively rare cases) crystallize out of a cooling magma.

Other minerals that can form in the context of saturated brines leading to the formation of evaporites include the **halides** and **sulfates**. The most common halide mineral is halite (NaCl). Another common mineral that can form during evaporation of seawater is gypsum, a calcium sulfate. The basic structural unit of sulfates is the sulfur-oxygen tetrahedron, i.e. an atom of sulfur surrounded by four atoms of oxygen (SO$_4^{2-}$). In the case of gypsum, water is also present in the crystal lattice and the chemical formula of gypsum is CaSO$_4$.2H$_2$O. When gypsum is buried under younger layers of sedimentary rocks, the increasing pressure and temperature squeeze out the water molecules and gypsum becomes anhydrite (CaSO$_4$).

Another common class of minerals includes the **sulfides**. In sulfides, the sulfide anion (S$^{2-}$) is bonded to metallic cations (e.g. Fe$^{2+}$, Mg$^{2+}$, Cu$^{2+}$). An example of sulfide is pyrite (FeS$_2$). Sulfides are important because they represent a major source of metals for the industry (metal ores). A geological setting in which sulfides commonly form is the ocean floor near mid-ocean ridges. Seawater percolates through cracks in the ocean crust where it is heated near spreading centers. The hot seawater reacts with the rock of the ocean crust (basalt) and dissolves iron, zinc, copper and sulfur. When this hot water returns to the ocean and meets the cold seawater, the sudden drop of temperature triggers the precipitation of sulfides which accumulate around the hot spring and builds a chimney called a **black smoker**. Black smokers are of huge interest not only because they are formed of metal sulfides but also because they shelter a diverse ecosystem whose primary source of energy is not sunlight (photosynthesis) but chemical reactions (chemosynthesis, e.g. $\text{CO}_2 + 4\text{H}_2\text{S} + \text{O}_2 \rightarrow \text{CH}_2\text{O} + 4\text{S} + 3\text{H}_2\text{O}$). Chemosynthesis can be performed by specialized bacteria and can sustain a diverse ecosystem in the deep ocean without any sunlight.

Another common class of minerals is called the **oxide**. In oxides the oxygen is bound to other atoms, usually metallic cations, through ionic bonds. One very common oxide and major source of iron for the industry (iron ore) is the iron oxide called hematite (Fe$_2$O$_3$). If the anion is not O$^{2-}$ but OH$^-$, the mineral is called a hydoxide. Another iron ore is the iron hydioxide goethite (FeO(OH)). Among the most famous rocks rich in iron oxide are the **banded iron formations** which are composed of thin layers of sediments rich in iron oxide. These rocks are usually Precambrian in age (older than 500 x 10$^6$ years) and they probably formed when the oxygen produced by photosynthesis reacted with the dissolved iron in the ocean. Their study is important to understand the history of the atmospheric oxygen of our planet (see the chapter on the evolution of life for further explanations).
**Native elements** are minerals composed of a single element. Sometimes they can be combined in alloys like the iron-nickel alloy of metallic meteorites. Other examples of native element minerals are native gold, native copper, graphite and diamond.

4.5. **Physical properties of minerals**

Minerals can be characterized by their physical properties. Some of the properties that can be used to characterize and identify minerals are listed below.

**Hardness**: harder minerals are able to scratch softer minerals. The hardest mineral is diamond which is able to scratch all the others. On the **Mohs scale** of hardness, diamond has a hardness of 10. The softest mineral is talc and has a hardness of 1. The hardness of a mineral depends on its structure and chemical bonds.

**Cleavage**: it is the tendency of minerals to split along specific planes. Chemical bonds along these cleavage planes tend to be weaker and thus break more easily.

**Fracture**: a mineral may be broken along surfaces that are not cleavage planes. These surfaces are called fractures. The quality (shape and texture) of the broken surface is one criteria used to identify minerals (e.g. the conchoidal fracture of quartz, the splintery fracture of fibrous minerals like asbestos).

**Luster**: it is the way the surface of a mineral reflects light. Different terms are used to qualify mineral luster (e.g. metallic, vitreous…).

**Color**: the color of a mineral is not always a good identification criterion because it may depend on the presence of tiny amounts of particular elements. For instance, quartz is colorless but amethyst, a variety of quartz with trace amounts of Fe$^{3+}$, is violet. Citrine, another variety of quartz, is yellow and can be obtain by heating amethyst. **Streak** refers to the color of the mark left by a mineral when it is scratched on an abrasive surface such as a plate of unglazed porcelain (e.g. the reddish brown streak of hematite).

**Density**: the density of a mineral (g/cm$^3$) depends on the way atoms are packed and on the mass of individual atoms. Diamond and graphite are both composed of carbon but diamond is denser because the atoms of carbon in diamond are packed more closely. Fe-olivine (Fe$_2$SiO$_4$) and Mg-olivine (Mg$_2$SiO$_4$) have a similar structure but Fe-olivine is denser than Mg-olivine because the atomic mass of iron is much greater than that of magnesium. The density of minerals is commonly expressed in terms of **specific gravity** which is the ratio of the weight of a volume of a mineral to the weight of an equal volume of water at a temperature and pressure of reference. The specific gravity of water is 1. The specific gravity of halite is 2.16 and that of pyrite is 5.

**Crystal habit**: it is the shape of an individual crystal or an aggregate of crystals of a given mineral. The crystal habit of a mineral reflects the internal arrangement of atoms as well as the speed and direction of crystal growth.
5. Introduction to rocks

5.1. What is a rock?

A rock is a naturally-occurring solid material usually composed of an aggregate of mineral matter. Two exceptions are coal and volcanic glass (obsidian). The former is composed in large parts of organic compounds which cannot be regarded as minerals. The latter is a solid glassy material in which the arrangement of atoms displays irregularities and therefore does not match the definition of a mineral. Properties of rocks that are commonly used to characterize them are their color, their texture (coarse-grained vs. fine-grained), and their mineralogical and chemical compositions.

5.2. Why study rocks?

We study rocks to learn about the history of the Earth. Rocks provide information about the origin and evolution of the terrestrial biosphere, lithosphere, atmosphere, and hydrosphere during the past 4.6 billion years. The study of rocks is not limited to our planet. By analyzing meteorites, rocks fallen from the sky, we can learn about the origin of the solar system. Samples of Moon rocks and soil have been collected during the Appolo missions and by the Russian Luna space crafts in the late 60s and early 70s. More recently, NASA rovers analyze Martian rocks to unravel the history of the red planet and to seek traces of present and past life. Rocks are important from an economical viewpoint as well because they contain valuable minerals. The industrial revolution during the first half of the 19th century would not have been possible without coal. Coal and other rocks rich in organic compounds are associated with the production of oil and natural gas. Rocks contain minerals that can be exploited if their concentration is sufficiently high. These rocks are called ores. An example of iron ore is the banded iron formation (BIF) already mentioned in the previous chapter. Rocks are also quarried for various purposes in the construction industry (e.g. concrete manufacturing, construction aggregates, building stones). Therefore, understanding how rocks form and how they are distributed is important for the mining industry. The study of rocks can help solving environmental issues. For instance, plans to store carbon dioxide within certain rock formations with the purpose to alleviate the current rise in global temperature are now being evaluated by experts. The storage of highly radioactive waste deep underground is also considered a relatively safe way to dispose of this dangerous material. Drinking water is a fundamental resource for all countries in the world. The presence or absence of groundwater reservoirs and their characteristics is determined by rock properties. The study of rocks is therefore essential for the localization, characterization, and sustainable exploitation of groundwater reservoirs.
5.3. Classification of rocks

Rocks can be classified into three families:

(1) **Igneous rocks** form by solidification of a cooling *magma* (molten rock). The size of mineral grains depends on the cooling rate (Fig. 1). Fine-grained igneous rocks are those which form near or at the surface of the Earth’s crust where magma cools more rapidly ( = *extrusive igneous rocks*). Volcanoes form where magma reaches the surface (more or less violently). If the cooling rate is extremely high (e.g. magma in contact with air or water), crystals may not even have time to form and *volcanic glass* is produced. Coarse-grained igneous rocks are those which form within the Earth’s crust ( = *intrusive igneous rocks*). They result from the slow cooling of magma within the crust. Bodies of magma present in the crust are called *magentic intrusions*.

![Figure 1](image)

**Figure 1**: Examples of igneous rocks (hand-size specimens, scale bar = 2 cm, and photomicrographs taken in cross-polarized light, scale bar = 1 mm). (A) Intrusive igneous rock (granite) with large mineral grains crystallized in a slowly cooling magma, (B) extrusive igneous rock (basalt) with tiny mineral grains crystallized in a rapidly cooling magma, and (C) extrusive igneous rock (basalt) with tiny mineral grains mixed with volcanic glass (in black on the photomicrograph) resulting from a very high cooling rate (note the shiny surface of the sample). Source: Imperial College Rock Library.

(2) **Sedimentary rocks** form by accumulation and subsequent *lithification* of *sediments* (fragments of preexisting rocks or elements of biological origin, Figs 2A-C) or by precipitation of minerals from an aqueous solution (Fig. 2D). Sedimentary rocks form in depressions of the Earth’s crust where sediments can accumulate. Most originate in the largest depressions, i.e. the ocean basins. Sediments can be fragments of igneous, sedimentary or metamorphic rocks, or fragments of individual minerals (e.g. quartz sand). Sediments can also be of biological origin, such as mollusk shells, coral skeletons, bones and plant remains. Sedimentary rocks formed by precipitation of minerals in saline lakes or embayments subject to evaporation are called *evaporites*.

![Figure 2](image)

**Figure 2**: Examples of sedimentary rocks (scale bar = 1 cm). (A) Conglomerate composed of rounded rock fragments (source: Imperial College Rock Library), (B) limestone composed of fragments of shells made of calcite (CaCO₃) (black arrows indicate shells of single-celled organisms called foraminifera), (C) coal (rock derived from an accumulation of plant debris, source: USGS), and (D) evaporite composed of layers of anhydrite (CaSO₄) (source: Garcia-Veigas et al., 2013).

---

*Lithification means solidification of a soft sediment (another word is *induration*). This process occurs primarily by *compaction* (from the overlying sediment load) and precipitation of minerals in the space between sediment grains (*cementation*).*
(3) **Metamorphic rocks** form by transformation of the chemical and/or mineralogical composition and/or texture* of a preexisting rock in a **solid state** due to changing conditions of temperature and/or pressure or due to interactions with hydrothermal fluids (Fig. 3). This process of rock transformation is called **metamorphism**. Rocks caught in subduction zones or between two colliding continents are subject to tremendous changes in temperature and pressure and can undergo various degrees of metamorphism. In addition, wherever rocks are in contact with magma, these rocks, if not melted, are transformed (metamorphosed) — in this case “cooked” — by increased temperature (**contact metamorphism**). In addition, water heated in the vicinity of magma circulates in the crust (**hydrothermal circulation**) and reacts with rocks, changing their chemical and mineralogical compositions (**metasomatism**). For instance, seawater which penetrates the oceanic crust through fractures near mid-ocean ridges is heated and leaches metals and sulfur from surrounding rocks. These elements precipitate as metal sulfides and oxides at hot springs, forming large chimneys rising from the sea floor (**black smokers**).

**Figure 3**: Examples of metamorphic rocks (hand-size specimens, scale bar = 2 cm, and photomicrographs taken in cross-polarized light, scale bar = 1 mm). (A) gneiss showing an alternation of darker and lighter layers with different mineralogical compositions (layers are perpendicular to the direction of pressure — indicated by the arrows — affecting the rock during metamorphism), (B) schist with elongated minerals oriented at right angle to the direction of pressure (foliation), and (C) marble (metamorphosed sedimentary rock, usually limestone. Source: Imperial College Rock Library.

The distribution of the different families of rocks is closely related to plate tectonics (Fig. 4). The formation of igneous rocks is linked to magma production and therefore to regions of the crust where rocks begin to melt. Rock melting takes place at mid-ocean ridges and subduction zones, at hotspots, and at the root of large mountain chains resulting from continental collision. Most sedimentary rocks are produced in ocean basins whose formation is controlled by plate tectonics (i.e. the opening of ocean basins through seafloor spreading). The distribution of metamorphic rocks too is closely linked to plate tectonics. Large variations of temperature and pressure affect rocks along convergent boundaries. Contact metamorphism occurs wherever rocks are in contact with magma. Metasomatism related to hydrothermal circulation occurs in the vicinity of magmatic intrusions associated with plate boundaries.

**Figure 4**: Idealized cross section of the Earth’s crust showing the regions of the crust where most igneous, sedimentary, and metamorphic rocks are produced. O-C SUBD. = ocean-continent subduction, HS = hot spot, O-O SUBD. = ocean-ocean subduction, MOR = mid-ocean ridges, CONT. COLL. = continental collision. Note that the hot spot location is not related to plate boundaries. Hot spots result from hot mantle material rising from great depths and producing volcanic activity at the surface (e.g. Hawaii).

* The texture of metamorphic rocks is determined by the size, shape and orientation of minerals.
Processes leading to the transformation of a rock of one family to a rock of another family can be described as a cycle called the **rock cycle** (Fig. 5). The rock cycle illustrates how each rock families can evolve from one another. The transformation of one rock family to another depends on plate tectonics and climate.

The control of plate tectonics on rock formation is clear when one looks at where the different types of rocks are being produced (see Fig. 5 and related text). Climate controls the production rate of sediments because the erosion of mountains depends on the amount of precipitation and other climatic factors (e.g. temperature). Hence, climate controls the production of sedimentary rocks. Plate tectonics also controls the production rate of sediments because plate tectonics controls the formation of mountains from which sediments are derived. Moreover, the long term evolution of Earth’s climate (over millions of years) is influenced by plate tectonics: (1) global volcanism along plate boundaries influences the concentration of CO$_2$ in the atmosphere; (2) the formation of large mountain chains along convergent plate boundaries (e.g. Himalayas) influences the concentration of CO$_2$ in the atmosphere through the process of silicate weathering (see chapter on sedimentary rocks).

**Figure 5.** The rock cycle. Each arrow represents a transformation process from one rock family to another. Each transformation process is represented by a specific color. Note that metamorphic rocks themselves can also be metamorphosed.
6. Igneous rocks

Igneous rocks form whenever molten rock cools and minerals crystallize to produce a solid rock. Hence the prerequisite for the formation of igneous rocks is the melting of rocks. Molten rock is called magma if it occurs within the lithosphere and becomes lava when the molten rock is extruded on Earth’s surface through volcanic activity.

Where does magma form?
Magma forms whenever the conditions of pressure and temperature cause a solid rock to start melting. This happens primarily in three major geological settings: (1) mid-ocean ridges, (2) subduction zones, and (3) hotspots. Details about the melting process in these three settings are presented in section 6.2.

6.1. Classification of igneous rocks

Igneous rocks are classified according to their texture (coarse-grained vs. fine-grained) and their chemical and mineralogical compositions. The texture and chemical or mineralogical compositions of an igneous rock can give us information on where and how the rock formed.

6.1.1. Classification based on rock texture

*Extrusive igneous rocks*
These rocks cool rapidly on or very near the Earth’s surface. The rapid cooling is due to the great difference in temperature between the hot molten rock and the relatively cool surrounding environment. These rocks are fine-grained since minerals crystallize too rapidly to form large crystals. A common example of an extrusive igneous rock is basalt. Basalt is a major component of the oceanic crust and produced in great abundance at mid-ocean ridges.

The extremely fast cooling of a lava extruded on Earth’s surface results in the production of *volcanic glass* in which atoms do not have time to form a regular crystal lattice.

When lava is ejected out of a volcano, pieces of lava cool and solidify in the air. The pieces of igneous rocks then fall on the ground by gravity. These pieces of igneous rocks are called *pyroclasts*. They may have various sizes, from volcanic *ash* (<2 mm) to large volcanic *bombs* (>6.4 cm). Pieces of intermediate size are called *pumice*. Explosive volcanic eruptions are sometimes characterized by a sudden and massive release of a dense mixture of extremely hot gas and ash rushing down the volcano’s flank at very high speed. This is called a *pyroclastic flow* and represents a major hazard for anyone being in the vicinity of the volcano when it happens.

*Intrusive igneous rocks*
These rocks cool slowly within the Earth’s crust. The slow cooling enables minerals to grow large. The texture of these rocks is coarse-grained. An example of an intrusive igneous rock is granite.
Granite is a major component of the continental crust.

Igneous rocks may sometimes consist of a combination of fine-grained and coarse-grained textures which tells us something interesting about their history. **Porphyry** is the name given to an igneous rock composed of large crystals (or phenocrysts) “floating” in a fine-grained matrix. This particular texture indicates that the rock started to cool slowly at some depth beneath the surface producing large crystals, and then began to cool faster as it approached the surface forming tiny crystals out of the remaining melt.

### 6.1.2. Classification based on chemical and mineralogical compositions

Igneous rocks enriched in silica (SiO$_2$) and silicates rich in Al, K, and Na (e.g. feldspar) are called **felsic** (from feldspar and silica). These rocks are characteristic of the continental crust (e.g. **granite**, Fig. 1)

Igneous rocks with a high proportion of silicates rich in Mg and Fe (e.g. olivine) are called **mafic** (from magnesium and ferric). These rocks are particularly abundant in the oceanic crust (e.g. **basalt**, Fig. 1).

Igneous rocks composed of a large proportion of silicate of Mg and Fe (e.g. olivine, (Fe, Mg)$_2$SiO$_4$) are called **ultramafic**. Ultramafic rocks are characteristic of the upper mantle. **Peridotite** is a rock composed mostly of olivine and pyroxene* and is believed to be a major constituent of the Earth’s mantle (Fig. 1).

Felsic and ultramafic rocks are the two end-members of a continuum of compositions. Igneous rocks may be more or less felsic, more or less mafic depending on how they formed. Igneous rocks with different chemical and mineralogical compositions have different properties. Felsic igneous rocks tend to be lighter than mafic igneous rocks. Felsic rocks have a lower density and start melting at a lower temperature. Magma with a felsic composition is more viscous than magma with a mafic composition (Fig. 1).

Note that the name given to igneous rocks (granite, basalt...) depends on both texture and composition. For instance, the name of a coarse-grained intrusive igneous rock with a felsic composition is **granite** whereas the name of the fine-grained extrusive rock with a similar composition is **rhyolite**. Also, **basalt** is the name given to a fine-grained extrusive igneous rock with a mafic composition but the intrusive equivalent is called **gabbro** (Fig. 1).

The figure below summarizes the main characteristics of extrusive and intrusive igneous rocks.

---

* Pyroxene is a silicate mineral formed of single chains of silicate tetrahedra. The pyroxene commonly occurring in peridotite is (Mg, Fe)SiO$_3$. The mineral with the formula MgSiO$_3$ is called enstatite. The mineral with the formula FeSiO$_3$ is called ferrosilite.
6.2. Processes of magma formation

6.2.1 Heat and pressure

Temperature increases with depth. This is called the geothermal gradient. On average, the geothermal gradient of the Earth’s crust is 30°C/km. As depth increases, temperature increases but pressure increases too which prevents rocks from melting. In “normal” conditions rocks of the crust and mantle cannot melt extensively. At mid-ocean ridges however where hot mantle rock rises to the surface, the geothermal gradient increases sharply and partial melting occurs at relatively shallow depth where the pressure becomes too low to maintain the whole rock at the solid state. The resulting melt is the source of the new oceanic lithosphere generated at mid-ocean ridges (see section 6.4). Since pressure drop plays such a crucial role in the melting process, it is called decomposition melting. Decompression melting occurs also beneath hotspots (e.g., Hawaii) where a mantle plume rises toward the surface. In this setting, the source of the plume may be as deep as the boundary between the mantle and the outer core, and therefore the rising mantle rock is even hotter than beneath mid-ocean ridges.

6.2.2. Influence of water

Besides mid-ocean ridges and hotspots, another geological setting where abundant magma is produced is the subduction zone. In subduction zones, an oceanic plate is sliding beneath another plate and plunges into the asthenospheric mantle. The subducting oceanic plate is relatively cold in comparison with the surrounding rocks. What is causing rock melting in this case? The answer is water and the process of magma formation is called water-induced melting. The subducting oceanic plate carries a lot of sediments and sedimentary rocks with relatively high water content. Water is present in the open space between sediment grains (pore space) and in the crystal lattice of clay minerals. Water molecules disrupt chemical bonds and lower the melting temperature of rocks.

* No melt is normally allowed except a tiny fraction of the upper asthenosphere (up to 1% of the rock volume only) directly below the moving tectonic plates.
6.2.3. Magma composition and style of volcanism

The composition of the rock from which magma is initially derived (**parent rock**) is important in determining magma composition. The magma may also gain components by melting the surrounding rocks during its ascension toward Earth’s surface. Magma composition also depends on temperature because different minerals melt at different temperatures (Fig. 1).

An important property of magma is its **viscosity**. A magma that is more viscous has more chance to remain stuck in the volcano’s vent. The pressure building up under this natural “cork” is particularly dangerous and may result in a violent explosive volcanic eruption.

The viscosity of magma is proportional to its silica content* and inversely proportional to its temperature (the cooler, the more viscous). The most violent volcanic eruptions are related to magmas that are both viscous and rich in volatiles (water, CO₂...). A volatile-rich, highly viscous magma is much more likely to give rise to an explosive eruption. A magma nearing the surface behaves like a soda drink when you unscrew the bottle cap. The volatiles contained in the magma form gas bubbles which expand due to the pressure drop. Decreasing the amount of volatiles dissolved in the magma make it also more viscous. A volatile-rich, highly viscous magma is likely to give rise to an explosive eruption due to the pressure that progressively builds up in the chimney and is suddenly released in the form of a violent explosion.

Hawaii is an oceanic hotspot where magma is derived from mantle rock. The magma has a mafic composition and therefore contains little silica. Moreover it is particularly hot because the parent rock originates from very deep in the mantle. As a result, Hawaiian volcanoes are characterized by a low-viscosity lava that can easily flow out and spread over large areas forming so-called shield volcanoes (volcanoes with low-angle flanks). The risk of explosive eruptions in Hawaii is very low.

Conversely, in continental hotspots (e.g. Yellowstone), continental rifts (e.g. East African Rift) and ocean-continent subduction zones (e.g. Japan), the magma may become more felsic as it penetrates the continental crust and melts the surrounding rocks. Felsic magma is highly viscous because it contains a lot more silica. The magma may also become enriched in volatiles. Consequently, the risk of explosive eruption in these settings is much greater!

6.2.4. Magma chambers

Magma has a lower density than the surrounding solid rock and therefore tends to migrate slowly toward the surface. It rises through cracks and by melting its way up. Magma tends to accumulate in regions of the crust called **magma chambers**. During its ascent, the magma composition may change by mixing with other magmas or by melting surrounding rocks.

6.2. Magma crystallization and the formation of igneous rocks

* Silica molecules (SiO₂) tend to form chains which greatly increase the viscosity of the magma. More felsic magmas are more viscous.
The process by which magma composition is modified during crystallization is called **magmatic differentiation**. Magmatic differentiation can lead to the formation of various igneous rocks from a single initial melt through the following processes:

1. Different minerals crystallize at different temperatures. The order at which minerals crystallize has been determined experimentally (see related slide showing the Bowen’s reaction series).
2. In a cooling magma, minerals crystallizing first tend to settle down first. This is called **crystal fractionation**. However, the crystal settling rate does not only depend on the timing of crystallization but also on the density and size of crystals and the viscosity of the remaining magma.

The result of magmatic differentiation is that a cooling magma produces a succession of igneous rocks whose chemical compositions vary from relatively more mafic to relatively more felsic.

Besides magmatic differentiation, other factors should be accounted for in order to explain the diversity and abundance of igneous rocks. Two of these factors are:

1. **Variability in parent magma composition**
   - At mid-ocean ridges and oceanic hotspots, the initial parent magma is derived from the melting of mantle rock and its composition is mafic (i.e. basaltic).
   - In subduction zones, the initial parent magma is derived from the melting of the oceanic crust (including various sedimentary rocks) and mantle rock. Its composition is mafic to intermediate (i.e. andesitic).
   - In the case of continental rifts, continental hotspots, and ocean-continent subduction zones in which continental crust is involved, magma may become more felsic (i.e. granitic) as it penetrates the continental crust and melt surrounding rocks.

2. **Mixing of different magmas**
   - Magmas that can mix with each other are said to be miscible. Crystallization of a magma resulting from the mixing of two miscible magmas produces rocks whose chemical compositions differ from those which would have been produced if each magma had crystallized separately. Some magmas however cannot mix and are called immiscible.

6.3. Forms of igneous intrusions

The path taken by the magma from its source to the surface can result in the formation of different geological structures described in figure 2.

6.4. Formation of new oceanic crust at mid-ocean ridges

The asthenosphere is convecting and asthenospheric mantle rock rises slowly toward the surface beneath mid-ocean ridges. The ascending hot mantle rock is composed of peridotite (olivine and lesser pyroxene and garnet). Decompression melting of the peridotite occurs at shallow depth. Up to 15% of the rock volume can melt. Magma accumulates in a magma chamber beneath the
mid-ocean ridge. Since pyroxene and garnet melt first, the composition of the resulting partial melt is not peridotitic but basaltic (enriched in silica and iron). The remaining solid, which has not melted, is enriched in olivine and forms a layer of ultramafic peridotite at the base of the oceanic crust. One part of the basaltic magma cools slowly in the magma chamber to form a layer of gabbro overlying the ultramafic peridotite. The other part of the basaltic magma rises to the surface and solidifies quickly near or on the seafloor to form a layer of basalt. As the two plates are pulled apart, the oceanic crust grows laterally by successive intrusions of basaltic magma which produce numerous dykes (see related slide for an illustration).

Figure 2: Cross section of a portion of the lithosphere through which magma is rising to the surface. Magma typically accumulates in the crust in large magma chambers (1). Pieces of surrounding rocks can be incorporated in the magma and change its chemical composition (2). The magma may start to crystallize upon cooling which also change the chemical composition of the remaining melt (3). Some of the magma leaves the chamber through fractures and may form large sheets, either parallel to rock layers — sill — (4) or intersecting rock layers — dyke — (5). A volcano is composed of a central vent (6) which may branch off and form a side vent (7). Lava is ejected from the vents and form lava flows (8). Other materials which may be ejected from vents are pyroclasts (ashes, volcanic bombs…) and gas (9).
7. Sedimentary rocks

Sediments and sedimentary rocks cover 90% of Earth’s surface. They are essential to our understanding of the origin and evolution of life because they are the rocks containing fossils. They are also the rocks from which coal, oil, and natural gas can be extracted, and are often associated with reservoirs of groundwater.

7.1. What are sediments?

Examples of sediments include beach sand, the gravels of riverbeds, and the fine particles desert dunes are made of. These are sediments that have not yet been transformed into sedimentary rocks. They are still soft, un lithified. We will see in section 7.3 how soft sediments become a hard, lithified rock.

Sediments can be classified into two broad categories: solid matter and ions in solution. Examples of solid sedimentary matter: (1) fragments of rocks and minerals, (2) the hard parts of organisms (biominerals, e.g. mollusk shells), and (3) organic matter (e.g. plant leaves).

Examples of ions in solution: Na⁺, K⁺, Cl⁻, Ca²⁺, CO₃²⁻...

7.2. How do sediments form?

7.2.1. Weathering

Sediments form by destruction of preexisting rocks. The destruction process is called rock weathering. Two types of weathering can be distinguished: (1) physical weathering and (2) chemical weathering.

(1) Physical weathering

Major physical weathering agents are wind (eolian weathering), water, and ice (glacial weathering). The dust carried by the wind abrades the surface of rocks. A good example of eolian weathering is the polished surface of stones in arid regions such as the Moroccan desert or Mars.

Water is a powerful weathering agent. The destructive action of waves along shorelines is a good example. Receding shorelines due to wave action are common around the world. Given enough time, rivers can cut through hard rocks and form V-shaped valleys with steep flanks. One of the most remarkable examples is the Grand Canyon in the US. Ice is also a powerful agent of rock destruction. Glaciers are like rivers of ice which can carve large U-shaped valleys. During the last ice age, the northern polar ice cap was much more extensive than today. Regions of North America and northern Europe that are now ice-free still bear the marks of glacial weathering. Ice can also destroy rocks through frost wedging. This weathering process results from the fact that the volume of water increases when it freezes. If liquid water fills the fractures in a rock and then freezes, ice expands and widens the fractures which can in turn break the rock. Plants may also participate in rock physical weathering through root wedging. Plant roots in the fractures of a rock
force them to open further as the plant grows which may lead to rock dislocation.

(2) Chemical weathering

Rainwater and CO₂ combine in soils to form the weak carbonic acid H₂CO₃. In contact with carbonate rocks, the carbonic acid dissolves CaCO₃. The products of this reaction, Ca²⁺, HCO₃⁻ and CO₃²⁻, are transported by rivers to the ocean. In the ocean, the ions Ca²⁺ and CO₃²⁻ can be used by marine organisms to build their calcareous hard parts (see related slide). Dissolution of CaCO₃ is responsible for the formation of caves and other dissolution features (karst). The process of CaCO₃ dissolution does not require the action of living organisms. However, the active dissolution of CaCO₃ by organisms happens as well. For example, some bivalves are capable of dissolving CaCO₃ and bore holes in which they live.

Like carbonate rocks, silicate minerals can react with acidic groundwater. The products of this weathering reaction include dissolved ions (e.g. Ca²⁺, Si⁴⁺, Fe²⁺, HCO₃⁻, CO₃²⁻, H⁺) and clays. These ions are transported by rivers to the ocean and some can be used by marine organisms to build hard parts made of CaCO₃ and SiO₂. The weathering of silicate rocks results in a net removal of CO₂ from the atmosphere over millions of years. This slow removal of CO₂ must be balanced by an equally slow input of CO₂. This input of CO₂ is provided by volcanoes and hot springs.

7.2.2. Erosion and transport

The removal of sediments from their source area is called erosion. Erosion is carried out by the same flowing agents causing rock physical weathering: wind, water, and ice. Sediments are then transported by these same agents to the site of their deposition. If there is no active removal, sediments may simply move away from their source area by gravity. The downslope movement of sediments by gravity is called mass wasting.

7.2.3. Sediment deposition

Solid sedimentary particles carried either by wind (eolian transport), water or ice (glacial transport) will at some point be deposited. Deposition of a sedimentary particle transported by wind or water occurs when the wind/water speed decreases and is no longer capable of carrying the particle in question.

In water, strong currents with a speed of >50 cm/s can carry all sizes of particles up to boulder-size (>25.6 cm). Moderately strong currents with a speed of 20-50 cm/s can carry sand (62.5 µm-2 mm) and smaller particles. Weak currents (<20 cm/s) can carry silts and clays (<2 mm). Note that we consider here only sediments consisting of fragments of rocks and minerals which generally have roughly the same density. We don’t consider material with a very low density like wood.

Sediments carried by ice deposit when the ice melts. The nature and morphology of these deposits is typical and can be recognized in the rock record. Glacial deposits are useful indicators of past climate.

In the case of mass wasting, the nature of sediments (e.g. size, shape) and slope steepness control deposition.
For ions in solution, deposition occurs either abiotically or biotically. An example of abiotic deposition is the precipitation of minerals in a saline lake by evaporation (see section 4.3). Biotic deposition of ions in solution is the process of biomineralization (see section 4.3). A good example is given by corals which use Ca\(^{2+}\) and CO\(_3\)\(^{2-}\) present in the ocean to produce their skeleton made of CaCO\(_3\).

7.3. How do sedimentary rocks form?

How soft sediments (say beach sand) are transformed into hard sedimentary rocks (in this case, sandstone)?

As soft sediments accumulate, they are progressively buried under younger layers of sediments. The set of physical and chemical changes which happen to soft sediments after their burial is called diagenesis. The result of diagenesis includes the transformation of soft sediments into hard rock. This transformation is called lithification. Lithification results from two main diagenetic processes:

1. **Compaction**: as sediments are buried deeper and deeper under younger sedimentary layers, the pressure and temperature increase. Sediments are squeezed, compacted. The space between sediment grains gets smaller.
2. **Cementation**: dissolved ions in groundwater precipitate in the open space between individual sediment grains. Common mineral cements include calcite (CaCO\(_3\)), hematite (Fe\(_2\)O\(_3\)), and quartz (SiO\(_2\)).

Compaction and cementation contribute to reduce the rock porosity. Porosity is a measure of the amount of open space (or pores) in a rock.

7.4. Properties of sediments and sedimentary rocks

The characteristics of sediment grains can provide information about the environment in which they formed. Fundamental properties of sediments and sedimentary rocks include:

**Grain size**: since sedimentary particles have generally roughly the same density, grain size gives a good indication of the strength of the current. The occurrence of coarse sand in a sediment indicates a stronger water current than a sediment composed entirely of mud.

**Sorting**: sorting is a measure of the variation in grain size within a sediment. If all sediment grains have approximately the same size, the sediment is well sorted. If the sediment is made of a mixture of small grains and large grains, the sediment is poorly sorted. This also provides information about the type of current which transported the sediment. If the strength of the water current is unchanging, the size of sediment grains deposited tends to be uniform. If the strength of the water current is variable, the size of grains deposited varies accordingly.
Grain morphology: during transport sediment grains are abraded and become rounded. Therefore the roundness of sediment grains provides information on the distance of transport. The greater the distance of transport, the greater the roundness.

7.5. Sedimentary basins and depositional environments

Large amounts of sediments are deposited in depressions of the Earth’s crust. Large depressions in which abundant sediments accumulate are called sedimentary basins. The largest basins are the ocean basins. Their size and morphology is controlled by plate tectonics. An example of a large continental sedimentary basin is the East African Rift which corresponds to a divergent boundary where two plates are pulled apart. The rift contains many lakes in which sediments can accumulate. In the future, the plates will keep moving away from each other. The continental lithosphere will stretch and get thinner. The rift will grow larger and deeper leaving more space for sediments to accumulate. The downward movement of the lithosphere as it is stretched and thinned is called subsidence. Subsidence is accentuated by the weight of sediments pushing down the lithospheric plate. Ultimately, seawater will flood the rift and seafloor spreading will begin, transforming the continental rift into an ocean basin. The two continental margins that are moving apart progressively cool as they are moving away from the source of heat. As it cools, the lithosphere gets denser and progressively sinks creating space for sediments to accumulate. This is a particular kind of subsidence called thermal subsidence. It enables the accumulations of thick piles of sediments along continental margins (continental shelf).

There exist many different kinds of depositional environments (deserts, lakes, rivers, delta, beach, organic reef, continental shelf, deep sea…) which belong to either one of the three following categories: continental, shoreline, and marine environments. Each depositional environment is characterized by a unique combination of physical, chemical, and biological processes. The nature of sediments deposited in these environments is determined by these processes. For example, the deep sea is not affected by strong, sustained currents like shallow-water environments. The deep-sea is therefore generally characterized by a fine-grained sedimentation (mostly mud). Deep-sea organisms are also different from those living in shallow waters. Let’s take another example: coral reefs. Coral reefs are associated to tropical and subtropical climate. They grow exclusively in shallow water and their rigid framework consists mainly of coral skeletons and crustose coralline algae. Although the organisms which compose them have evolved, organic reefs have been around for almost as long as the Earth has existed (e.g. stromatolites, see section 7.6.2).

Useful indicators of past depositional environments are sedimentary structures. Although many sedimentary structures exist, only the most common are described herein. When one looks at an outcrop of sedimentary rocks, the most conspicuous sedimentary structure is often the bedding or stratification plane (see related slide for illustration). Sedimentary beds can be a few cm thick to
several m thick. Bedding results from changes in sedimentation affecting grain size and/or sediment composition. The equivalent of bedding at the mm scale is called \textit{lamination}. Bedding is useful to study tectonic deformation (folds and faults) because sediment layers are usually deposited flat and horizontal. Other remarkable sedimentary structures are \textit{ripple marks} (see related slide for illustration). Ripple marks are a common feature of sandy beaches. The geometry of ripple marks preserved in sedimentary rocks provides information on the transport agent (wave vs. wind, unidirectional vs. bidirectional, current direction). \textit{Burrows} are also sedimentary structures. They represent one example of sedimentary structures resulting from the activity of living organisms. Other biological sedimentary structures include traces left at the surface of sediments by benthic organisms (benthic means living on the seafloor).

In conclusion, geologists are able to reconstruct past sedimentary environments based on the characteristics of sedimentary rocks. When you have a succession of sedimentary rocks of different ages at an outcrop or in a drill core, it becomes possible to study environmental changes over geological time scales. The fossil content of the rock can also be used to understand the evolution of life.

7.6. Types of sedimentary rocks

7.6.1. Siliciclastic sedimentary rocks

\textbf{Siliciclastic sedimentary rocks} are composed of fragments of rocks (lithic fragments) and minerals. The name “siliciclastic” comes from the fact that the composition of these rocks is dominated by silicate minerals (e.g. quartz). Siliciclastic rocks are classified according to grain size (coarse = \textit{conglomerate}, medium = \textit{sandstone}, fine = \textit{shale}) or grain composition (e.g. arkose = sandstone containing at least 25% of a mineral called feldspar).

7.6.2. Biochemical sedimentary rocks

\textbf{Biochemical sedimentary rocks} are usually formed by accumulation of the hard parts of marine organisms (e.g. bivalve shells, coral skeletons...). A beach deposit consisting of fragments of bivalve shells (\textit{bioclasts}) is an example. Coral reef framework is another example. Bivalve shells and coral skeletons are made of CaCO$_3$. Another major biomineral is silica (SiO$_2$). Marine microorganisms such as diatoms and radiolarians secrete a tiny shell made of silica that can accumulate on the deep seafloor.

The constituents of biochemical sedimentary rocks are generally formed by the direct biological precipitation of CaCO$_3$. But biological activity can also induce precipitation of CaCO$_3$ indirectly. Photosynthetic microbial mats called stromatolites are a good example. The equilibrium between dissolved calcium and carbonate ions in the sea and the mineral calcium carbonate can be written as follows: $\text{Ca}^{2+} + 2\text{HCO}_3^- \leftrightarrow \text{CaCO}_3 + \text{CO}_2 + \text{H}_2\text{O}$. The reaction of photosynthesis is $6\text{H}_2\text{O} +$
6CO₂ + sunlight → C₆H₁₂O₆ + 6O₂. Photosynthesis consumes CO₂ whereas calcification releases CO₂ in the environment. Consequently, the photosynthetic activity of stromatolites tends to promote calcification by shifting the equilibrium state of the carbonate system to the right (Ca²⁺ + 2HCO₃⁻ → CaCO₃ + CO₂ + H₂O). Stromatolites can form rigid pinnacle-like structures and these structures owe their rigidity to the early cementation induced by photosynthesis (see related slide for illustration).

7.6.3. Chemical sedimentary rocks

**Chemical sedimentary rocks** are formed by precipitation of minerals which does not involve biological activity. For example, minerals precipitating at the bottom of a saline lake subject to intense evaporation. The accumulation of minerals in this kind of setting produces a rock called *evaporite*. One of the most common constituent of evaporites is NaCl (halite or table salt).

7.6.4. Organic sedimentary rocks

**Organic sedimentary rocks** are formed by accumulation of organic matter (e.g. plants). Usually, organic matter is quickly decomposed by bacterial activity. However, in particular settings which are characterized by a high rate of accumulation of organic matter and low levels of oxygen, organic matter can be preserved and accumulate.

An example of such accumulation is *peat*. Peat forms in wetlands by accumulation of plants. Low oxygen concentrations prevent the complete decomposition of organic matter. As peat is buried under younger sediments, the increase in pressure transforms peat into *lignite* (compressed peat). At greater depth, pressure and temperature increases further and lignite is converted into *coal*. Coal is a complex mixture of mineral matter and organic carbon compounds. Increasing further the pressure and temperature transforms coal into *anthracite*. Anthracite has a higher carbon content than coal. Coal and anthracite have a different chemical composition than peat and lignite. Heat and pressure have modified their composition and texture. They are not strictly speaking sedimentary rocks anymore but metamorphic rocks (see next chapter).
8. Metamorphic rocks

In chapter 7, we talked about the transformation of soft sediments into hard sedimentary rocks by compaction and cementation. These processes occur once the sediments are buried under new sedimentary layers. Thick accumulation of sediments and sedimentary rocks is made possible by a process called **subsidence** (see chapter 7). As depth increases, temperature and pressure increase too. At a certain depth, temperature and pressure are so high that it may affect the mineralogy, texture or chemical composition of rocks. Rocks undergo a process called **metamorphism**. Most metamorphic rocks are produced in the middle and lower crust at a depth of 10 to 30 km. Plate tectonics leads to metamorphism where shallow rocks are forced down to great depths during the process of mountain building at convergent plate boundaries (i.e. orogeny). But metamorphism can also occur at or near Earth’s surface where rocks are in contact with hot magma (i.e. contact metamorphism, see section 8.4). The **metamorphic grade** of a rock refers to the pressure and temperature to which the rock has been subjected. Metamorphic rocks formed at relatively low pressure and low temperature have been subjected to a **low-grade metamorphism**. Those produced at high pressure and high temperature result from a **high-grade metamorphism**.

It is important to understand that rocks remain **solid** during the process of metamorphism. If rocks melt, the resulting magma will produce an igneous rock which is another domain of the rock cycle.

Rock metamorphism is driven by three factors that are described below: (1) temperature, (2) pressure, and (3) heated fluids (hydrothermal fluids).

8.1. Temperature

The average **geothermal gradient** of the Earth’s crust is 30°C/km. A rock buried 15 km below the surface is subjected to a temperature of 450°C. At such temperature, a rock re-crystallizes. Preexisting minerals grow larger and new ones may even form. Different minerals are stable at different temperatures. For example, chlorite is a sheet-like silicate commonly forming during low-grade metamorphism. Garnet, another silicate mineral, is characteristic of a higher-grade of metamorphism. The temperature at which one mineral assemblage changes into another has been determined experimentally. Therefore, by studying the mineral composition of a metamorphic rock, geologists can determine the temperature at which the rock has formed. Metamorphic mineral assemblages are natural **geothermometers**.

The depth at which metamorphism occurs depends on the geothermal gradient. The geothermal gradient can be more or less steep according to the geological setting. An old, stable continental crust has a geothermal gradient that is lower than average. On the other hand, regions where the crust is stretched and thinned (e.g. Basin and Range in US, East African Rift) are characterized by a geothermal gradient higher than average.
8.2. Pressure

Due to the weight of overlying rocks, pressure increases with depth in Earth’s crust at a rate of 300-400 bar/km. The pressure at 15 km is thus approximately 5000 bar. For comparison, the atmospheric pressure at sea level on Earth is 1 bar. In the ocean, the pressure in the deepest trenches about 10,000 m below the surface is approximately 1000 bar. The pressure resulting from the weight of overlying rocks is called confining pressure. Crustal rocks are squeezed in all directions by this confining pressure which increases with depth. There exists another kind of pressure: the directed pressure caused by tectonic forces. For example, there are tremendous compressional forces affecting crustal rocks at convergent boundaries. The shape and orientation of minerals formed during metamorphism are greatly influenced by tectonic forces. Elongated and platy crystals tend to orientate perpendicularly to the direction of compression. The parallel alignment of platy crystals in a metamorphic rock produces a distinct texture called foliation (see section 8.5). Crystal orientation in a metamorphic rock is therefore useful to reconstruct the tectonic history of a region. Moreover, like for temperature, different minerals are stable at different pressures. Geologists can use the mineral assemblages they recognize in metamorphic rocks to determine the pressure at which they formed. Metamorphic mineral assemblages are natural geobarometers.

8.3. Hydrothermal fluids

Another factor that can lead to modifications of the mineralogical and chemical composition of a rock is the presence of heated fluids (hydrothermal fluids) in the crust. Even in dry rocks, water can be present in the crystal structure of some minerals (e.g. clay minerals). Increasing pressure eventually squeezes out water molecules. Heated water can react with surrounding rocks and dissolve components like CO$_2$, SiO$_2$, Na, K, S, Cu, and Zn, changing the initial composition of the rocks. New minerals can also precipitate from a hydrothermal fluid as it penetrates different regions of the crust. The transformation of rocks’ mineralogical and chemical composition in contact with hydrothermal fluids is called metasomatism.

An example of such transformation is related to the formation of black smokers, already mentioned in chapter 4. Seawater penetrating the oceanic crust near mid-ocean ridges is heated and reacts with the surrounding basalt. The heated water leaches elements such as sulfur and various metals from the rock. Meanwhile, the mineralogical composition of the basalt changes and new minerals form (e.g. chlorite). Due to the presence of these new minerals, the color of the rock changes too and becomes green (hence the name given to this type of rock: greenstone). When the heated water returns to the ocean floor and cools suddenly in contact with the cold deep ocean water, metallic sulfides precipitate and form large chimneys from which seeps a blackish mixture of hot water and mineral particles (hence the name black smoker).
8.4. Types of metamorphism

**Regional metamorphism**: metamorphism which affects large regions of the crust at convergent boundaries where rocks are subjected to high temperature and high pressure. Rocks that are carried to huge depths in subduction zones can undergo high to ultra-high pressure metamorphism. Metamorphic rocks formed under very high pressure are called **eclogites**. They are rarely brought back to the surface and therefore rarely seen at outcrops. They contain minerals indicative of extremely high pressures, such as **coesite** (a high-pressure polymorph of quartz) or microdiamonds which form under 40,000 bar at a depth of 120 km!

**Contact metamorphism**: metamorphism which affects rocks in contact of a magmatic intrusion within the crust or in contact with lava at the surface. In this case rocks are subjected to high temperature but not necessarily to high pressure.

**Seafloor metamorphism**: metamorphism which affects rocks of the oceanic crust near mid-ocean ridges in contact with hydrothermal fluids (metasomatism, see explanations in section 8.3. of this chapter)

**Burial metamorphism**: metamorphism which affect sedimentary rocks as they are buried under an increasingly thick pile of sediments. This metamorphism typically starts at depths ranging between 6 to 10 km where rocks undergo a pressure of about 3000 bar and temperatures are comprised between 100 and 200°C.

**Shock metamorphism**: metamorphism which is caused by a meteorite impact. The kinetic energy of a meteorite is released in the form of heat and shock waves. Temperature can be high enough to melt rocks at the impact site. The rock is pulverized and small molten pieces cool in the air and form small glassy bead-like objects called **tektites**. The presence of tektites in sedimentary rock layers is used as one of the criteria to recognize a meteorite impact in the rock record. The shock wave can modify the mineralogy of the rocks at the impact site. For examples, the crystal structure of quartz can be modified to produce coesite. The presence of high-pressure minerals like coesite or microdiamonds can also be used as supporting evidence for a meteorite impact in the rock record.

8.5. Metamorphic textures

Metamorphism guides the size, shape, and orientation of crystals and has a profound influence on the texture of rocks. A common texture observed in metamorphic rocks is **foliation**, already mentioned in section 8.2 of this chapter. Foliation results from the preferential orientation of platy crystals which tend to align in the direction perpendicular to compressional forces. Minerals forming platy crystals and common in metamorphic rocks are **micas**. The foliation plane is a plane
along which the rock splits more easily (metamorphic cleavage). For example, slate is a low-grade foliated metamorphic rock composed of small platy crystals whose preferential orientation enables the rock to be split easily into flat sheets. Due to this property, slates are commonly used to make floor and roof tiles.

Based on their texture, two main families of metamorphic rocks can be distinguished: foliated metamorphic rocks and non-foliated (granoblastic) metamorphic rocks.

8.5.1. Foliated metamorphic rocks

Foliation is produced when metamorphic rocks are subjected to directed pressure. As the grade of regional metamorphism increases, foliation becomes more pronounced and the size of crystals increases. Foliated metamorphic rocks include the following common rock types, classified here from a lower (top of the list) to a higher (bottom of the list) grade of metamorphism (see slides for illustrations):

- slate
- phyllite
- schist
- gneiss
- migmatite

Foliation becomes more conspicuous as the grade of metamorphism increases from slate to phyllite to schist. In gneiss, minerals are segregated into distinct lighter and darker bands. In migmatites, banding becomes more diffuse as temperature increases and the rock approaches its melting temperature.

8.5.2. Granoblastic (non-foliated) metamorphic rocks

Granoblastic rocks are metamorphic rocks which have not been subjected to directed pressure and consequently do not display any foliation. One example of such rock type is greenstone. Greenstones form in the context of seafloor metamorphism as explained in section 8.3. Another example is marble. Marbles form by recrystallization of sedimentary rocks composed primarily of CaCO₃ (limestones) subjected to high pressure and high temperature. Quartzite, another granoblastic metamorphic rock, is the name given to a sandstone rich in quartz grains that has undergone high pressure and high temperature. Note that in metamorphosed sedimentary rocks such as marble and quartzite, fossils are destroyed during the process of recrystallization.

8.5.3. Porphyroblastic texture

The term porphyroblastic is used to describe a rock composed of large crystals (porphyroblasts) “floating” in a much finer-grained matrix. The interpretation of this texture is that the large crystals have formed over a broad range of pressure and temperature. The small crystals in the matrix on the other hand are not stable over such a broad range of pressure and temperature. They have had to recrystallize repeatedly as pressure and temperature increased which did not give them enough time to grow large.
8.6. Metamorphic rocks: a tool to study the history of crustal rocks

8.6.1. Index minerals
As explained in section 8.1 and 8.2, mineral assemblages in metamorphic rocks can be used as a kind of geothermobarometer. Of particular interest are so-called index minerals which are stable in a relatively limited range of pressure and temperature. Mapping the distribution of index minerals in a region where metamorphic rocks are exposed to the surface enables to delineate zones of the crust which have experienced different grades of metamorphism. The lines delimiting these zones are called isograds.
Ancient mountain chains are often characterized by extensive belts of different metamorphic grades (with different index minerals) parallel to their axis with isograds following the main deformation features of the rocks (folds and faults). Therefore, mapping metamorphic zones is useful to recognize ancient convergent plate boundaries. In addition, the mineralogy of metamorphic rocks gives access to the conditions of pressure and temperature existing when they formed. The next step is to identify the precise geological setting to which these metamorphic rocks are related.

8.6.2. Metamorphic facies
The mineralogy of metamorphic rocks does not only depend on pressure and temperature but also on the composition of the parent rocks. As an example, let’s consider a shale (mainly clay minerals and quartz) and a basalt (mainly feldspars and pyroxene). Let’s raise the pressure and temperature so that both rocks experience the same intermediate-grade metamorphism. The mineralogy of the metamorphosed shale and that of the metamorphosed basalt will overlap (presence of garnet in both rocks) but will not be the same. Both metamorphic rocks however belong to the same metamorphic facies (in this case the facies called amphibolite) because they are characteristic of a specific domain of pressure and temperature typically found deep under mountain chains. Another example is the metamorphic facies called greenschist which includes a variety of rock compositions characterized by a high abundance of green minerals such as chlorite and which typically forms by low-grade metamorphism under mountain chains. Rocks belonging to the blueschist facies on the other hand are characterized by the presence of the blue mineral glaucophane and indicative of conditions of moderate temperature and high pressure typically experienced by rocks of the oceanic crust caught in a subduction zone. There are of course other metamorphic facies characteristic of other tectonic settings which are not mentioned here.
By subjecting rocks of various compositions to various conditions of pressure and temperature in the lab, geologists can match the compositions of metamorphic rocks from a surveyed area to specific domains of pressure and temperature (hence metamorphic facies) which provides valuable information about the geological setting in which these rocks formed (e.g. subduction zone, mountain belt, contact metamorphism).
8.6.3. The pressure-temperature paths of metamorphic rocks

The history of pressure and temperature of a given metamorphic rock is called its pressure-temperature path or \textit{P-T path}. The P-T path consists of two segments: a prograde and a retrograde segment. The \textit{prograde segment} reflects the increase in pressure and temperature since the rock began to experience metamorphism. The \textit{retrograde segment} reflects the decrease in pressure and temperature when the rock is brought back to the surface through the process of \textit{exhumation} (see section below). Rocks from different tectonic settings have different P-T paths (see slides for illustration). It is therefore very useful to know the P-T path of a metamorphic rock.

In order to reconstruct the P-T path of a metamorphic rock, geologists target large \textit{porphyroblasts} (see section 8.5.3). One example is \textit{garnet} which grows over a broad range of temperature and pressure. The key point here is that the chemical composition of garnet evolves as pressure and temperature increase. Experiments conducted in the lab have shown exactly how the composition changes with pressure and temperature. Hence, by analyzing changes in the abundance of specific elements along the transect connecting the center of a porphyroblast of garnet (oldest age) to the rim of the same crystal (youngest age), it is possible to reconstruct the evolution of pressure and temperature during the crystal’s growth.

8.7. Exhumation process

Metamorphic rocks can sometimes be brought back to the surface (luckily for geologists who can then study these rocks). This process is called \textit{exhumation}.

One hypothesis suggests that exhumation results from the combined effect of \textit{tectonic uplift} and \textit{erosion}. At convergent boundaries, rocks at the core of mountain chains are uplifted by tectonic forces. Meanwhile the surface of the mountain chain is affected by erosion which removes rocks from the top. The result is that deep-seated rocks become progressively closer to the surface as they are pushed upward by tectonic forces and the rocks above are removed by erosion. Since erosion is controlled by climate, this theory implies that \textit{climate} and \textit{tectonic uplift} control the rate of exhumation.
9. The age of rocks

9.1. Absolute vs. relative ages

The issue of time is central to earth science. Geologists want to know how old rocks are and how long it takes for geological processes to occur. Until the 20th century and the discovery of radioactivity, geologists could only tell whether a rock was younger or older than another rock. In other words, they could only determine the relative age of rocks. Knowing the number of years elapsed since a rock has formed, in other words knowing the absolute age of the rock, could only be achieved after the development of radiometric dating in the 20th century.

9.2. The relative age of rocks

9.2.1. Stratigraphic principles

Stratigraphy is the study of sedimentary layers (or strata). The founder of stratigraphy is a Danish scientist named Nicolas Steno (1638-1686). He is also known for his study of the fossil Glossopetrae which he correctly identified as fossilized shark teeth, a notable achievement at a time when the true nature of fossils was poorly understood. His interest for geology led him to study the sedimentary rocks of northern Italy (where he lived) and to enounce two fundamental rules of stratigraphy:

1. Principle of original horizontality
Sediments are deposited horizontally (or nearly so) by gravity. This implies that sedimentary layers which are folded or faulted have been subjected to deformations after their deposition.

2. Principle of superposition
New sedimentary layers form on top of older layers. Therefore in an undeformed succession of sedimentary layers, layers at the bottom are older than layers at the top.

The principle of superposition is of course very useful to determine the relative age of sedimentary layers. When studying a section of sedimentary rocks, this principle can be used to reconstruct the chronological order of the layers providing that we understand the deformation history of the rocks in question.

What if we have two sections that are very distant from each other? How can we know that the rocks of one section are younger or older than the rocks of another section located far away? To solve this problem, one way is to study the fossil content of sedimentary layers.

9.2.2. Biostratigraphy

* Due to folding, a succession of sedimentary layers can become up-side down and the chronological order inverted. If the deformation is not noticed, the chronological reconstruction will be completely wrong!
Biostratigraphy is the branch of stratigraphy which uses fossils to correlate sedimentary rocks and find out their relative ages. The technique relies on the fact that life evolves. Some species go extinct while new species appear. As a consequence, fossil assemblages of different geological intervals are not the same. The geological record is characterized by fossil assemblages whose species composition is changing through time. This is the principle of faunal succession. For example, a fossil of Australopithecus (one ancient genus of hominids — the family of primate to which we belong — who appeared around 4 million years ago and went extinct 2 million years later) cannot be found in the same layer as a fossil of dinosaur because dinosaurs went extinct 65 million years ago (61 million years before the first Australopithecus appeared!). Some species are more useful than others when it comes to determine the relative ages of sedimentary rocks. Widespread species which have existed for a limited time and are characteristic of specific geological periods are excellent biostratigraphic markers. These species are called index species.

NB: the concept of the evolution of life was accepted before Charles Darwin (1809-1882) proposed his theory of evolution by natural selection. The roots of biostratigraphy are to be found in the 18th century when scientists noticed that sedimentary rocks of different ages were characterized by different fossils. What Darwin discovered is a mechanism explaining how new species arise and therefore how evolution can take place.

9.2.3. Unconformities: gaps in the stratigraphic record
The stratigraphic record is not continuous. There are gaps which can result from prolonged lack of deposition or from the erosion of preexisting sedimentary layers. The surfaces corresponding to these gaps are called unconformities.
In many cases unconformities result from an episode of erosion. Sedimentary layers formed in a lake or in the sea can be eroded during a long period of emergence. How do sedimentary layers initially formed under water become emerged? There are two important mechanisms that can cause long period of emergence: sea level fall (1) and upward ground movement or uplift (2).

(1) During ice ages (glacial periods), more ice accumulates on landmasses and sea level falls dramatically. During the maximum of the last ice age about 20,000 years ago (Last Glacial Maximum = LGM), the sea level was 125 m lower than today! The continental shelves stood above sea level and sediments were exposed to the erosional action of the wind and the rain. New sedimentary layers formed only when the sea re-flooded the continental shelves during the current warm period. The surface which separates these new layers from the layers deposited before the LGM is an unconformity and represents a time gap of thousands of years.

(2) Regions of the crust near convergent plate boundaries experience tremendous compressional forces which can cause folding and bring portions of the crust at higher elevations, a process
called tectonic uplift*. Uplifted sedimentary layers can be raised above sea level where they are exposed to erosional processes. Once compression stops, an episode of subsidence (downward movement of the crust, see section 7.5) may create a depression in which new sediments can accumulate. The surface between the older folded sedimentary layers and the new sediments is an unconformity which may encompass a time lapse of millions of years.

There are three types of unconformities:

**Disconformity**: the sedimentary layers below and above the unconformity are both undeformed and horizontal. An example of disconformity is the unconformity corresponding to the episode of sea level fall presented above [see (1)].

**Angular unconformity**: the sedimentary layers below the unconformity are folded whereas the sedimentary layers above the unconformity are undeformed and horizontal. An example of angular unconformity is the unconformity related to the episode of compressional deformation followed by erosion, subsidence, and deposition of new sediments presented above [see (2)].

**Nonconformity**: unconformity between sedimentary rocks and metamorphic or igneous rocks. This type of unconformity may represent a very extensive time gap of tens of millions of years or more.

### 9.2.4. Cross-cutting relationships

The geometrical relationships between sedimentary rock formations, igneous intrusions (e.g. dykes — sheet-like intrusion intersecting rock layers — ), faults, and unconformities can be used to reconstruct the chronological order in which these geological features formed. The rule is simple: *younger geological structures cut older ones*. For example, a dyke intersecting sedimentary layers must be younger than these layers or a fault cutting a dyke must be younger than the dyke.

### 9.2.5. The geological time scale

Geologists of the 19th century used the principle of superposition and biostratigraphy to divide the geological record into successive intervals characterized by distinct fossil assemblages. This approach led to the construction of the **geologic time scale**. The basic subdivisions of the geologic time scale (from longer to shorter time units) are **eras** (e.g. Cenozoic), **periods** (e.g. Quaternary), and **epochs** (e.g. the current epoch called Holocene).

The boundary between geologic periods is characterized by abrupt changes in fossil assemblages. Several of these boundaries correspond to mass extinctions. A mass extinction event represents a relatively short geologic time span (a few million years or less) during which a large proportion of the total number of species living on the Earth (e.g. 75% or more) disappears (see chapter on the evolution of life).

* Note that uplift does not necessarily imply folding. Rocks can be uplifted without undergoing much folding.
9.3. The absolute age of rocks

When the geologic time scale was established in the 19th century, geologists did not know the duration of each period. Nobody knew precisely how old the Earth was. This issue was important not only for science but also for religion because of the strongly held religious belief that the Earth could not be older than a few thousand years based on the rigorous interpretation of religious texts. The solution came only after the discovery of radioactivity in 1896 by a French physicist named Henri Becquerel. At the beginning of the 20th century, the physicist Ernest Rutherford proposed a technique to date rocks based on radioactive decay. The method of radiometric dating was born. Based on this method, an American geochemist named Clair C. Patterson calculated in the 1950s an age for the Earth of 4.56 billion years!

The basic principle of radiometric dating is simple. Elements consist of different isotopes with nuclei composed of the same number of protons but with different numbers of neutrons. Radioactive isotopes are isotopes which spontaneously disintegrate (decay) into a different element. In a rock containing a certain amount of a given radioactive isotope, the concentration of the element produced by radioactive decay increases with time and can be used as a natural clock to determine the age of the rock. This can work only if the rate of radioactive decay is known and does not vary in time.

The rate at which a radioactive isotope (parent atom) disintegrates into another element (daughter atom) is expressed by its half-life. The half-life of a radioactive isotope is the time its takes for half of the initial amount to decay into daughter atoms (Figs. 1 & 2).

![Graph showing the decay of radioactive isotopes](image)
Different isotopes have different half-lives. For example, $^{14}{\text{C}}$ (carbon-14) disintegrates into $^{14}{\text{N}}$ and has a half life of 5730 years. Another example is $^{87}{\text{Rb}}$ which disintegrates into $^{87}{\text{Sr}}$ and has a half-life of 49 billion years. Therefore if you want to measure the age of rocks that are hundreds of millions of years old or more, you can use the rubidium-strontium system but not the radioactive isotope of carbon. The $^{14}\text{C}$ method is used to measure the age of much younger materials. For example, you can use $^{14}\text{C}$ to measure the age of corals which were growing thousands of years ago.

*The half-life of a particular radioactive isotope is constant.* It is not affected by changes in physico-chemical conditions, such as variations in temperature and pressure. This is of fundamental importance because if the half-life of radioactive isotopes varied in time, radiometric dating would not be possible.

In order to understand how the age of a rock can be measured, let’s take the example of the rubidium-strontium system. In this case, the parent atom is $^{87}\text{Rb}$ and the daughter atom is $^{87}\text{Sr}$. Let’s say we collected a sample of igneous rock and we want to know its age. The age of the sample in this case is the time elapsed since it crystallized from a cooling magma. The crystals forming in the cooling magma have trapped a certain amount of parent atoms $[^{87}\text{Rb}]_{t=0}$. First let’s suppose that there is no daughter atoms $^{87}\text{Sr}$ incorporated in the crystals when they formed (unrealistic!). We can express the amount of $^{87}\text{Rb}$ and $^{87}\text{Sr}$ present in our sample as a function of time using the following relationship:

$$[^{87}\text{Sr}]_{t} = [^{87}\text{Rb}]_{t=0} \left[1 - (1/2)^t\right]$$

$$[^{87}\text{Rb}]_{t} = [^{87}\text{Rb}]_{t=0} \left(1/2^t\right)$$

$$\frac{[^{87}\text{Sr}]_{t}}{[^{87}\text{Rb}]_{t}} = \frac{[^{87}\text{Rb}]_{t=0} \left[1 - (1/2)^t\right]}{[^{87}\text{Rb}]_{t=0} \left(1/2^t\right)} = \frac{1 - (1/2)^t}{1/2^t} = \frac{1}{1/2^t} - \frac{1/2^t}{1/2^t} = 2^t - 1$$

$$[^{87}\text{Sr}]_{t} = (2^t - 1) \left[^{87}\text{Rb}\right]_{t} \quad \text{(A)}$$
Equation (A) represents a straight line with a slope equal to \((2^t - 1)\) (Fig. 3). Therefore measuring the amount of \(^{87}\text{Sr}\) and \(^{87}\text{Rb}\) in our sample would enable us to calculate the age of the rock.

In reality, there is always a certain amount of daughter atoms \(^{87}\text{Sr}\) which is incorporated in the crystals when the rock forms. In such case, equation (A) becomes (Fig. 4):

\[
[^{87}\text{Sr}]_t = (2^t - 1) \left[^{87}\text{Rb}\right]_t + [^{87}\text{Sr}]_{t=0}
\]  

(B)

In this case, measuring the amount of parent atoms and daughter atoms in our sample is not enough to calculate the age of the rock because we don’t know how much daughter atoms has been trapped initially in the rock (Fig. 5). Each minerals of our rock sample can incorporate any amount of \(^{87}\text{Rb}\) and \(^{87}\text{Sr}\) at the time of crystallization.
The solution to solve this problem is to consider the ratio of \(^{87}\)Sr to a stable isotope of the same element which has the same properties. Here the stable isotope in question is \(^{86}\)Sr. Since \(^{87}\)Sr and \(^{86}\)Sr have the same behavior during physico-chemical reactions, they will incorporate different minerals always in the same proportion. Let’s say we have 1000 atoms of \(^{87}\)Sr and 1200 atoms of \(^{86}\)Sr initially present in the magma. The table below shows their distribution in three different minerals. The initial ratio \(^{87}\)Sr/\(^{86}\)Sr is independent of the amount of \(^{87}\)Sr trapped in the minerals when they crystallized.

<table>
<thead>
<tr>
<th>Mineral 1 (M1)</th>
<th>Mineral 2 (M2)</th>
<th>Mineral 3 (M3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>([^{87}\text{Sr}]_{t=0})</td>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td>([^{86}\text{Sr}]_{t=0})</td>
<td>600</td>
<td>120</td>
</tr>
<tr>
<td>([^{87}\text{Sr}] / [^{86}\text{Sr}]_{t=0})</td>
<td>0.83</td>
<td>0.83</td>
</tr>
</tbody>
</table>

If we divide equation (B) by \([^{86}\text{Sr}]_{t}\), we obtain the following relationship:

\[
\frac{[^{87}\text{Sr}]}{[^{86}\text{Sr}]}_t = (2^t - 1) \frac{[^{87}\text{Rb}]}{[^{86}\text{Sr}]}_t + \frac{[^{87}\text{Sr}]}{[^{86}\text{Sr}]}_{t=0} \tag{C}
\]

We can thus determine the age of our sample by measuring the ratio of \(^{87}\)Sr and \(^{87}\)Rb to \(^{86}\)Sr in several minerals in order to obtain a straight line from which we can derive the age \(t\) (Fig. 6). In

* Note that the isotopic composition of the magma should be uniform when crystallization occurs.
addition to the age, we can also find out the initial ratio of $^{87}\text{Sr}$ to $^{86}\text{Sr}$.

The slope of the line defined by the three minerals in the graph above is a function of the time elapsed since crystallization. This method is valid only if the minerals have not exchanged elements with the surrounding environment after they crystallized. Each mineral must have remained a closed system\(^*\). For metamorphic rocks, the age obtained corresponds to the time since the last phase of crystallization. We can also measure the age of biominerals like coral skeletons or algal crusts made of CaCO\(_3\). In this case, we measure the time elapsed since the organism secreted the mineral assuming that the system remained closed afterward.

\(^*\) Note that there are techniques dealing with open systems.
10. Continents: structure and history

We saw in chapter 3 that the Earth’s outer shell—or lithosphere—is broken into plates moving relative to one another (plate tectonic). The lithosphere comprises the crust and the uppermost part of the upper mantle. The crust is subdivided into an oceanic and a continental crust. The former is heavier, thinner, and lies beneath the ocean floor which represents 2/3 of the Earth’s surface. The latter is lighter, thicker, and makes up the continents. The oceanic lithosphere is produced at mid-ocean ridges and destroyed along subduction zones. Therefore, the oldest oceanic lithosphere present beneath the ocean floor is only 200,000,000 years old, which represents a relatively small fraction of Earth’s 4.6-billion years history. On the other hand, continental rocks can be much older because they are too light to be subducted. Hence, only continental rocks can be used to study the evolution of our planets in the distant past (billions of years ago).

The edge of a continent facing a subduction zone is called an active margin. A good example is the west coast of South America. The edge of a continent that does not coincide with a convergent plate boundary is called a passive margin. The Atlantic Ocean is bordered by passive continental margins.

10.1. Structure of continents

The surface of continents can be subdivided into several broad regions which were formed at different times and by different tectonic processes. These broad regions are referred to as tectonic provinces:

**Shields** lie at the heart of continents and are their oldest components. Shields are composed of Precambrian crystalline rocks (metamorphic and igneous rocks). These rocks may have been deformed during the Precambrian but remained undeformed during the Phanerozoic. Examples of shields are the Canadian Shield, the African Shield, and the Australian Shield.

**Platforms** consist of sedimentary layers overlying the Precambrian basement and displaying relatively minor deformations (e.g. North American Platform, Siberian Platform). Shields and platforms consist of the most stable parts of continents called cratons. Cratons have not been significantly affected by mountain building episodes (orogenies) during the Phanerozoic. However, they may have experienced slow vertical movements due to mantle convection beneath the lithosphere or to the growth and retreat of ice sheets (see epeirogeny, section 10.3). Large areas can be uplifted due to heating by mantle upwelling beneath the lithosphere\(^*\). More locally, magmatic intrusions migrating upward can warp the overlying rock layers. Circular regions of the lithosphere which have been uplifted are called structural domes (e.g. Cincinnati Arch, USA; Richat Dome, Mauritania). Rock layers in a structural dome dip away from its center. On the other hand, regions of the lithosphere which cools and contracts are affected by subsidence and are

\(^*\) Note that vigorous upwelling may cause crustal extension which in turn may lead to rifting and the formation of a sedimentary basin. An ocean may even form if ocean-floor spreading is initiated and maintained (see the Wilson cycle, section 10.2.3).
called thermal subsidence basins where abundant sediments can accumulate (e.g. Michigan Basin, USA; Amadeus Basin, Australia). Unlike domes, rock layers in a thermal subsidence basin dip toward its center. Intracratonic basins are economically important because they contain large reservoirs of coal, oil, and natural gas.

Phanerozoic orogens are mountain chains situated along the margins of cratons and produced by continental collision or ocean-continent subduction. The Appalachian mountains (eastern USA) and the Ardennes (central Europe) are examples of past orogens related to mountain building episodes that are no longer active. On the other hand, the Alpine-Himalayan orogen and the Andes are examples of ongoing mountain building episodes (see slides for more details).

Regions of extended crust are regions where the last episode of tectonic deformation is an extension. Passive margins are regions of extended crust because they are formed after rifting and the subsequent opening of an ocean basin (e.g. the passive continental margins bordering the Atlantic Ocean).

10.2. Growth of continents

Two main processes contribute to the growth of continents: one that involves vertical transport of matter (magmatic addition) and one that involves horizontal transport of matter (accretion).

10.2.1. Magmatic addition

Abundant magma is produced at subduction zones through water-induced melting of the subducting oceanic plate and the mantle directly above it (see chapter 6). After being produced, the magma rises toward the surface and either crystallizes in a magmatic chamber or is spewed out during volcanic eruptions. In Japan, the occurrence of many granitic intrusions and active volcanoes attests of the importance of magmatic addition for the formation of the Japanese Archipelago.

10.2.2. Accretion

A continent can also grow larger when a portion of the neighboring plate merges with the continent. Accreted blocks of tens to hundreds of kilometers in size are called terrains (or terranes). The geology and age of these accreted terrains may be very different from those of the adjacent continental rocks. Japan is formed of a juxtaposition of accreted terrains parallel to the subduction zone and bounded by major faults. The age of terrains decreases toward the trench. They are composed of portions of the oceanic crust as well as volcanic and sedimentary rocks carried by the subducting plate. An example of accreted material observed in Japan is the siliceous sedimentary rock called chert (see chapter 7). It is a layered rock composed of tiny shells made of silica which have accumulated on the deep ocean floor. Another example of accreted material in Japan is reef limestone. A prominent example is the Akiyoshi Terrain in Yamaguchi Prefecture which is

* Note the analogy with the thermal subsidence experienced by passive margins and permitting the thick accumulation of sediments on continental shelves.
composed of atoll reefs ranging in age from Early Carboniferous to Middle Permian. Continental growth by accretion does not only happen in subduction zone. Probably the most obvious example of continental growth by accretion is when two continental landmasses merge during continental collision (e.g. collision between India and Eurasia forming the Himalayan mountain chain). Less obvious but nonetheless of fundamental importance is the accretion of blocks moving along a transform fault. This process has shaped the west coast of North America. For example, the terrain named Wrangelia now forming part of Alaska and western Canada originated from the southern hemisphere and has travelled 5000 km north to finally merge with North America.

10.2.3. **The Wilson cycle**

The margins of cratons have experienced several episodes of mountain building throughout Earth’s history. Meanwhile ocean basins have opened and closed repeatedly. Since the surface area of the Earth is constant, plate divergence must be compensated by convergence. Divergence is linked to the opening of ocean basins whereas convergence leads to the closure of ocean basins and continental collision. An idealized picture of the evolution of continents and ocean basins is given by the *Wilson cycle*. Once during the Phanerozoic all the continents merged into a single supercontinent called *Pangaea* about 250 million years ago. Pangaea was not the first supercontinent and will not be the last. The Wilson cycle begins with the break up of a supercontinent through rifting and the opening of an ocean basin followed by a subduction leading to continental collision and ultimately to the assembly of a new supercontinent. Rifting of the new supercontinent marks the beginning of the next cycle. It takes several 100s of millions of years to break-up and then reassemble a supercontinent.

10.3. **Vertical movements of the continental lithosphere**

Continental collision results in a great deal of crustal deformation (faults and folds), metamorphism, and crustal thickening leading to the formation of magma at the root of mountain chains. But the continental lithosphere can also be affected by vertical movements with little or no deformation. These vertical movements of the continental lithosphere are referred to as *epeirogeny*. There are two main causes of epeirogenic movements: (1) glacial rebound and (2) the heating/cooling of the lithosphere.

10.3.1. **Glacial rebound**

The weight of large glaciers can bend the continental lithosphere. When these glaciers melt, the lithosphere rises progressively. High-latitude regions of the northern hemisphere like Scandinavia are now being uplifted following the melting of the thick ice sheet that covered these regions during the last ice age (110,000-12,000 years ago).
10.3.2. Heating/cooling of the lithosphere

Upwelling of mantle rock beneath the continental lithosphere causes uplift. When heated, the lithosphere expands and its density decreases which makes it “float” higher on the asthenosphere. Moreover, the pressure of the rising mantle rock may also push the lithosphere upward. Uplift of the lithosphere may lead to the formation of structural domes, already mentioned in section 10.1. The Basin and Range Province in western USA is a result of the heating of the lithosphere by mantle upwelling inducing uplift and crustal extension accompanied by faulting.

Conversely, cooling of the continental lithosphere causes subsidence. This process leads to the formation of thermal subsidence basins where abundant sediments can accumulate (e.g. Michigan Basin, passive margins of continents).

Thermal subsidence and uplift are processes occurring over millions of years. Conversely, glacial rebound is faster and occurs over thousands of years.

10.4. Origin of the resistance of cratons

Cratons are composed of very old crustal rocks. Evidence for continental crust 4 billion years old has been found in Canada. Cratons are very stable. Cratons may contain rocks which have remained undeformed since the Archean era (3.9-2.5 billion years ago). On the other hand, the margins of these stable cratons have been affected by multiple orogenic episodes. How can we explain the extraordinary resistance and stability of the cratons?

The continental lithosphere is light and cannot be subducted. On the contrary, rocks of the oceanic crust are constantly being recycled in the mantle. Hence, the rocks beneath the ocean floor are less than 200 million years old whereas those of the continental crust can be billions of years old. But how have cratonic rocks resisted tectonic deformation for so long? The answer probably lies in the nature of the lithosphere beneath the cratons. Cratons are characterized by a particularly thick lithosphere with a rigid lithospheric mantle that acts like a bumper and prevent cratons from being crushed during collisions. The thick and rigid lithospheric mantle beneath cratons is referred to as the cratonic keel. There is much speculation concerning its structure and composition. The low geothermal gradient beneath cratons suggests that cratonic keels are colder than the surrounding mantle rocks. If their composition was the same as the mantle material around, they would tend to sink because of their lower temperature, hence their greater density. One hypothesis suggests that cratonic keels are composed of a material that is less dense than the mantle around. Evidence supporting this hypothesis comes from a rock called kimberlite. This rock is formed during a violent volcanic eruption which brings material lying deep in the lithosphere up to the surface. These kimberlites are found within cratons and contain fragments of the cratonic keel (together with diamonds). The composition of these fragments is that of a peridotite. The rock is similar to “standard” mantle peridotite except it contains less iron and less garnet and consequently has a relatively lower density.
11. Origin and evolution of life (part I)

Geologists learn about the evolution of life on our planet by studying rock layers containing fossils and other traces of biological activity. The task is difficult because the geological record is incomplete. Incompleteness comes from unconformities in the stratigraphic record (see chapter 9) and processes destroying fossils (i.e. metamorphism, see chapter 8). An analogy can be made between the reconstruction of life history based on fossils and reading a book with missing pages. The study of past life based on fossils is called paleontology. Paleontologists assign fossils to taxonomic groups (groups defined on the basis of shared morphological characteristics), determine how and where fossil organisms lived (paleoecology), and study their evolutionary relationships.

A look at the fossil record clearly shows that the biological landscape of our planet has gone through considerable changes over time (think about the giant dragonflies of the Carboniferous or the dinosaurs of the Jurassic!). The fossil record is punctuated by extinction events and the emergence of new species. Life is not static but evolves through time!

A breakthrough in our understanding of life history came in the 19th century when Charles Darwin explained evolution in light of his theory of natural selection (a theory proposed independently by the British naturalist Alfred Russel Wallace at roughly the same time). Evolution is the process by which organisms descend from other preexisting forms of life through modifications. These modifications are the result of random genetic mutations affecting specific traits (e.g. morphology, physiology, behavior). If they are beneficial for the organism and increase the chance of survival and reproduction, they will have a greater chance to be passed on from one generation to the next (hence the term “natural selection”). Over many generations, these modifications can change the genetic make-up of a population and lead to the emergence of a new species (speciation). Advantageous traits that are inherited are called adaptations. Evolution by natural selection is a slow process. During rapid environmental changes, populations may not have time to adapt and entire species may face extinction. The biosphere has been affected several times by mass extinctions which have wiped out a large proportion of species living on the planet. However, these major disturbances were followed each time by a phase of rapid diversification as we will see in chapter 12.

In light of Darwin’s theory life can be represented by a tree, an evolutionary tree, in which each branch represents a group of organisms (Fig. 1). The branches at the very top of the tree represent extant groups. Moving down these branches means traveling back in time. The junction between two or more branches represents the time when new groups have evolved from a common ancestor. If we follow the branch of this common ancestor, we see that it too shares a common ancestor with another group. In this tree, we can find a common ancestor even for extant groups that are very different from each other, such as eagles and slugs, if we travel back in time far enough. Darwin himself surmised that there must exist at the base of the tree one “primordial life form” from which all the others have descended (Fig. 1).
An evolutionary tree can be reconstructed on the basis of the morphological traits shared by organisms. For example, we, humans, have a lot more characteristics in common with chimpanzees than with sea urchins. Therefore, our branch must be closer to that of chimpanzees and the branch of our most recent common ancestor must be relatively close to the tree top. On the other hand, the branch of sea urchins is far more distant from us and the most recent ancestor we share with them must be found further back in time, thus farther from the tree top. Comparing the morphological characteristics of extant groups of organisms does not tell us how their ancestors looked like, nor when they branched off to form new groups, and it says nothing about their relationships with extinct species. To answer these questions, scientists turn to the fossil record. For example, paleontologists have shown that the earliest birds probably emerged in the Late Jurassic and descended from dinosaurs.

Recently, molecular studies have provided new insight in the evolutionary relationships between extant groups of organisms (a discipline called phylogenetics). The more similar two groups are from the viewpoint of their DNA or proteins, the closer they are in the evolutionary tree and the more recent their common ancestor is. Phylogenetic studies can even provide information on the timing of divergence between two groups of organisms. In conclusion, data from fossils and molecular studies are complementary and used together to unravel the evolutionary history of life.

11.1. The origin of life

11.1.1. What is life?

In order to answer the question of the origin of life, we must first define what life is. What distinguish a living thing from a non-living thing? Let’s consider the most basic unit of life: the cell. The simplest organisms are composed of a single cell devoid of nucleus: the prokaryotic cell. Prokaryotes comprise two domains of life: bacteria and archaea. They share with all other forms of life a capacity of growth, reproduction, and evolution. They must take up matter from their environment to grow (carbon and other elements, i.e. nutrients) and need a source of energy. The capacity of evolution is related to the transmission of genes from one generation to another. In all organisms, DNA is the molecule carrying the genetic information. DNA must be able to make copies of itself in order to pass on the genetic information to the next generation (DNA replication). A living cell is also characterized by a permeable membrane which isolates and protects the

---

* Dividing life in groups on the basis of shared characteristics is the job of taxonomists. The father of taxonomy is Carl Linnaeus (1707-1778) who laid down the foundation of the current system of naming and classifying organisms. The modern classification comprises the following subdivisions: domain, kingdom, phylum, class, order, family, genus, and species.

** The technique is referred to as “molecular clock”. The rate of modifications in DNA and proteins is roughly constant among groups of organisms. Therefore the differences between the DNA or proteins among distinct groups of organisms can be used to infer the amount of time elapsed since they diverged from a common ancestor.
genetic material and which helps maintain an internal equilibrium (i.e. homeostasis).

There are many ways to define life. For instance, the physicist Stephen Hawking proposes the following definition: “One can define life to be an ordered system that can sustain itself against the tendency to disorder, and can reproduce itself”. From this perspective, even a computer virus turns out to be alive!

Whatever the definition of life we choose, we should now ask ourselves how the origin of life can be studied. The geological record can tell us a lot about primitive life forms but it cannot teach us exactly how life came to be in the first place. To investigate the origin of life, we have to turn to lab experiments.

11.1.2. The lab approach

There are three basic ingredients needed for the emergence of life as we know it: (1) water, (2) organic molecules, and (3) energy. Water is an essential constituent of life. The body weight of some organisms is composed of 90% of water! Water is also an excellent solvent and can transport elements involved in chemical reactions. Organic molecules are composed of a backbone of carbon atoms linked to other elements, primarily H, O, N, P and S*. Energy is needed to build the large organic molecules needed to sustain life (e.g. proteins, RNA, DNA).

Here are three important steps toward the formation of a living cell:

1. Formation of simple organic molecules

The first breakthrough in our understanding of the origin of life came in 1953 with the Miller-Urey experiment (Fig. 2). Its aim was to reproduce the conditions existing on the early Earth and see whether organic molecules could be produced. The experiment used water vapor and a mixture of gases (CH₄, NH₃, H₂, H₂O) to simulate the primitive atmosphere** and an electric arc as a source of energy to mimic the effect of lightning. Under these conditions, numerous organic molecules could be synthesized, including a variety of amino acids. Other experiments using the same approach but with different mixtures of chemical compounds succeeded in synthesizing nucleotides, key components of RNA and DNA.

Another hypothesis suggests that organic compounds may have been carried to Earth by meteorites and comets. A chemical analysis of the Murchinson meteorite showed that it contains various organic molecules, including amino acids.

In conclusion, large amounts of organic molecules must have existed on the early Earth. However, the absence of an ozone layer* implies that they were exposed to intense UV radiations. Life must have formed in an environment protected from these radiations, such as the underside of rocks or at some depth under water.

---

* C, H, O, N, P, S: the 6 elements essential for life are also among the most common elements in the universe!

** Note that there was no free oxygen in the primitive atmosphere. Oxygen was produced later by photosynthesis.

* Ozone (O₃) is produced in the upper atmosphere by the reaction O₂ + O → O₃ in which O comes from the dissociation of O₂ by solar radiation
Formation of complex organic molecules capable of self-replication (RNA, DNA)
The next step toward the emergence of life is the formation of complex organic molecules (polymers like RNA consisting of a chain of nucleotides). RNA is similar to DNA but instead of being double-stranded and helical, it is formed of a single strand of nucleotides. Biochemists have shown that clay minerals can promote the assembly of nucleotides into strands of RNA. DNA is like a blueprint that carries the information needed for building proteins. However, DNA does not directly communicate with the protein factory. Instead, RNA is used to copy the information stored in DNA (transcription) and carry it to the protein factory (Fig. 3A). It is also involved in other tasks like delivering the building units of proteins (i.e., amino acids). During cell division, DNA replicates itself so that the genetic information is passed on to the daughter cells (Fig. 3B). DNA replication requires protein enzymes to split the double-stranded DNA molecule and help connecting each strand with free nucleotides. The information for building these proteins is encoded in DNA. Life as we know it is therefore based on complex interactions between DNA, RNA and proteins. How to bridge the gap between the “prebiotic soup” and a complex DNA-based system? One possible scenario involves RNA. RNA is a versatile molecule that can encode genetic information and be the catalyst of its own replication. RNA could potentially have sustained a simple form of life before evolving into the more complex DNA-based system. This hypothesis is referred to as the RNA-world hypothesis.

Formation of membrane-bound vesicles capable of self-reproduction
It has been shown that lipid vesicles can form abiotically and even divide spontaneously under certain conditions (Fig. 4). The first protocell may have been composed of such lipid vesicles enclosing an information-carrying molecule such as RNA. Spontaneous division of these vesicles must have occurred in parallel with the self-replication of information-carrying molecules in order to pass on the genetic code to the next generation of protocells. Many of these protocells must have failed but those which bore advantageous modifications carried on their division and evolved by natural selection toward life as we know it.

11.2.2. Evidence from the geological record
The geological record can provide information on the environmental conditions existing on the early Earth. From the oldest Precambrian fossils, we can learn about the first primitive forms of life.
and the timing of biological evolution.

The Earth is 4.56 Ga old. Life would have had a hard time to form before 3.9 Ga because of repeated and violent meteorite impacts and a surface temperature above 100°C (remember the Heavy Bombardment in chapter 2). The first persistent oceans must have formed after this period around 3.8 Ga ago. The first evidence for the deposition of sediments under water comes from a 3.8 Ga rock formation in Greenland. Moreover, there was not or very little O₂ in the atmosphere and the first cellular life forms must have been anaerobic prokaryotes. Volcanic eruptions released gases such as CO₂, H₂S, SO₂, and N₂. Other gases that may have been present in the atmosphere are CH₄, NH₃, and H₂. The relative proportion of these gases is still debated.

Evidence for biological activity in the geological record can be classified in 4 categories:

① **Morphological evidence**
Microfossils and microbial mats can be preserved in sedimentary rocks which have not been metamorphosed.

② **Biomineralization**
There are two types of biomineralization: direct and indirect. An example of direct biomineralization is the precipitation of CaCO₃ by organisms to build a shell or skeleton (e.g. bivalves, reef corals). Minerals can also be a byproduct of biological activity. An example of indirect biomineralization is the precipitation of CaCO₃ in relation with bacterial growth in Stromatolites (see section 7.6.2.). In conclusion, the presence of specific minerals in rocks can indicate a possible link with biological activity.

③ **Stable isotope fractionation**
Stable isotope fractionation is the process by which stable isotopes are segregated on the basis of their mass during a physicochemical or biochemical process. For example, the evaporation of water involves the fractionation of the stable isotopes of oxygen. The ratio $^{18}$O/$^{16}$O of the water molecules in water vapor tends to be lower than in the liquid from which it evaporated. That is because the water molecules escaping from the liquid and turning into a gas tend to be enriched in $^{16}$O (the "light" isotope of oxygen). Similarly, biological processes produce isotope fractionation. Photosynthesis, for example, produces organic compounds enriched in $^{12}$C relative to $^{13}$C. Therefore, the isotopic analysis of sedimentary rocks can yield important clues regarding the occurrence and nature of biological processes.

④ **Biomarkers**
Biomarkers are large organic molecules produced by living organisms (e.g. polypeptids, triterpenes, and steranes). Their presence in sedimentary rocks provides a strong evidence for

---

* Ga stands for Giga Annum which means Billion of Years in Latin.
biological activity. Moreover, biomarkers may also give an indication on the type of organisms from which they derived. For example, the organic compound called algaenan is exclusively produced by green algae.

Evidence provided by stable isotopes suggests that biological activity existed as early as 3.8 Ga. The first morphological evidence comes from the remains of microbial mats (stromatolites) discovered in Australia in a 3.5 Ga old rock formation (still somewhat controversial though). There are multiple lines of evidence suggesting a diverse microbial life between 3.4-3.2 Ga. In conclusion, we can infer from the geological record that life may have formed as early as 3.8 Ga and was certainly present by 3.4 Ga.
12. Origin and evolution of life (part II)

The earliest form of life was probably a prokaryote. Unlike the eukaryotic cell, the prokaryotic cell has a more simple structure with no nucleus and no membrane-bound organelle (organelle = specialized subunit of the cell). This early prokaryote must have been able to survive without oxygen (anaerobic prokaryote). Anaerobic metabolism is carried out today by some species of bacteria and archaea.

In the text below, I review some of the main stages of the evolution of life and the evidence from the geological record.

12.1. The Great Oxygenation Event (GOE)

The free oxygen (O₂) present in the atmosphere today was probably produced initially by photosynthetic prokaryotes similar to modern cyanobacteria (the same microorganisms behind the formation of stromatolites, see chapter 7). The first unambiguous occurrence of cyanobacteria in the geological record is in 2.15 Ga old rocks from Canada. Fossilized stromatolites and microfossils with an age of 3.5 Ga suggests an earlier origin (but controversial, Fig. 1B). Geological and molecular evidence suggests that cyanobacteria evolved after various other types of bacteria, including anoxygenic photoautotrophs (organisms performing photosynthesis which do not produce oxygen).

Figure 1: 3.5 Ga old microfossils (A) and stromatolites (B) from Western Australia (sources: A. Schopf, 1993; B. Van Krakendonk, 2006)

The oxygen produced by photosynthesis did not accumulate directly in the atmosphere. Oxygen, a powerful oxidizing agent, first reacted with various elements. Massive amounts of iron oxide (Fe₂O₃) precipitated in the oceans when Fe²⁺ (ferrous ions)-rich water mixed with O₂. These "mass-rusting" events led to the formation of thick accumulations of red stratified sedimentary rocks rich in iron oxide called Banded Iron Formations (BIFs, Fig. 2A). Once the oxidation process was completed, oxygen started to accumulate in the atmosphere and gave rise to the Great Oxygenation Event (GOE) around 2.4-2.3 Ga ago. The occurrence of fluvial deposits older than 2.4 Ga containing sediments composed of mineral grains that would not be stable under oxidizing conditions confirms the lack of free oxygen in the atmosphere prior to the GOE (Fig. 2B).

The GOE had a major impact on life. First, oxygen being toxic for anaerobic bacteria, the GOE probably caused the first major mass extinction event. Second, the GOE led to a major innovation: the aerobic metabolism. Cellular respiration using O₂ as an electron acceptor produces much more energy than anaerobic respiration and was an essential step to sustain the energetic needs of more complex forms of life. Third, the adaptation of large prokaryotic cells to the toxicity of
oxygen is probably the mechanism behind the emergence of eukaryotes.

Figure 2: Banded Iron Formation (A) and 2.7 Ga old Witwatersrand conglomerate (South Africa) containing Pyrite flakes and uraninite (B) (sources: A. Encyclopedia Britannica, B. Natural Museum of Humbolt State University)

12.2. The eukaryotic cell

Another important step in the evolution of life is the development of a new way to obtain carbon by feeding directly on biological matter: predation. This innovation requires the cell to have a flexible, non-rigid membrane to engulf its prey. The eukaryotic cell may have evolved from the ingestion by large anaerobic bacteria of small aerobic bacteria which later became a structural part of the larger cells. The small aerobic bacteria may have survived inside the larger cells by using organic molecules produced by their host. The larger anaerobic bacteria on the other hand may have benefited from the ability of aerobic bacteria to use $O_2$ and produce energy through aerobic respiration. These aerobic bacteria may have evolved into the mitochondria of eukaryotic cells, organelles used for cellular respiration. In a similar manner, photosynthetic bacteria may have been ingested by larger cells and evolved into chloroplasts, organelles of plant and algal eukaryotic cells which carry out photosynthesis. This scenario, which implies the evolution from a predator-prey to a host-symbiont relationship, is called the endosymbiont hypothesis (proposed by Lynn Margulis in 1966, Fig. 3).

The oldest fossils of eukaryotes are found in a 1.8 Ga old rock formation in China (Fig. 4). Biomarkers indicate the existence of eukaryotes as early as 2.7 Ga. A recent molecular study suggests that the oldest ancestors of extant eukaryotes may have emerged between 2.5 and 1.6 Ga ago, and then quickly diversified.

Figure 3: The Endosymbiont Hypothesis (source: Biology, Life on Earth, 9th edition)

Figure 4: 1.4 Gyr eukaryote from Canada (A) and 1.8 Gyr eukaryote from China (B) (sources: A. Butterfield, 2005, B. Lamb et al., 2009)

12.3. Multicellularity

Once predation evolved, larger single-celled organisms may have had an advantage over smaller ones. For a cell, it is easier to engulf a prey which is comparatively smaller. Larger cells have also less chance to be ingested by others. However, larger single-celled organisms require more
exchange of matter through the cell membrane (intake of O\textsubscript{2} and nutrients, waste removal). Since the surface/volume ratio decreases as the cell becomes larger, there is a limit to the size of a single-celled organism (Fig. 5). Beyond such limit, the surface of the cell is not large enough to permit enough exchange to sustain cellular metabolism. The solution to grow larger without slowing down metabolic reactions is to adopt a \textit{multicellular} body. Besides greater predation efficiency and better protection against predator cells, multicellularity enables the evolution of specialized cells and body parts. For example, a multicellular alga can have roots to anchor itself on the seafloor and leaves to harvest the Sun’s energy for photosynthesis.

Figure 5: Surface area and volume relationships (source: Biology, Life on Earth, 9\textsuperscript{th} edition)

In the fossil record, the oldest undisputed occurrence of a multicellular organism with a clear affinity with an extant group is the red algae \textit{Bangiomorpha pubescens} found in a 1.2 Ga old rock formation in Canada (Fig. 6A). An older fossils of unknown affinity and possibly derived from a multicellular organism is \textit{Grypania spiralis} (Fig. 6B). The oldest specimen of \textit{G. spiralis} is from a 1.85 Ga old rock formation in US.

Multicellular organisms evolved shortly after the GOE once aerobic respiration provided the energy needed to sustain more complex forms of life.

Figure 6: 1.2 Ga multicellular red algae \textit{Bangiomorpha pubescens} from arctic Canada (A) and 1.6 Ga problematic \textit{Grypania spiralis} from India (sources: A. Butterfield, 2000, B. Butterfield, 2009)

12.4. The Cambrian radiation of life

A sharp increase in animal diversity is recorded in the fossil record at the end of the Precambrian 600-550 million years ago. The first phase of animal diversification produced what is known as the \textit{Ediacara biota} (Fig. 7). Ediacaran animals were the first large and complex forms of life. They had a bilateral or radial symmetry and were mostly soft-bodied. Their taxonomic relationship with extant groups is still debated. They disappeared at the onset of the Cambrian period. Ediacara marks the transition between two distinct biospheres: the Precambrian biosphere dominated by single-celled organisms (“the age of bacteria”) and the Cambrian biosphere comprising all known animal phyla.
Cambrian life is characterized by a major innovation: the shell or carapace. All known animal phyla were already present at the beginning of the Cambrian. This explosion of life forms is known as the **Cambrian radiation of life**. The diversification of life and the emergence of modern phyla had already begun in the Precambrian. For example, the first traces of Porifera (sponges) and Cnidaria (anemones, corals...) are found in Precambrian rocks. The explosion of diversity around the Precambrian-Cambrian boundary gave rise to an astonishing variety of body plans and architectures, many of which seem to have no equivalent today and may represent evolutionary “dead-ends” (Fig. 8).

A combination of factors is probably responsible for the Cambrian radiation of life. Here are several factors which may have played an important role:

1. **Environmental factors**
   - End of the Proterozoic glaciations
   - Rise in atmospheric O$_2$* providing an energetic advantage
   - Change in ocean chemistry facilitating biomineralization

2. **Genetic factors**
   - Emergence of key developmental genes resulting in new evolutionary possibilities

3. **Ecological factors**
   - Adaptations related to a predator-prey relationship. For example, preys evolved traits to avoid being caught by predators, such as a shell or a carapace, a pelagic** mode of life, or body parts which improved their mobility. Predators on the other hand evolved traits to help them catch preys, such as improved sensory organs (e.g. the eye). The evolutionary interdependence between different groups of organisms is referred to as coevolution (distinct groups of organisms can influence each other’s evolution).

**12.5. The colonization of land**

---

* Some scientists suggest that this second “oxygenation event” may have been triggered by the break-up of the supercontinent Rodinia. The break-up of Rodinia created more extensive continental margins on which large amounts of organic matter could be buried and escaped decomposition (C$_6$H$_5$O$_6$ $\rightarrow$ 6CO$_2$ + 6H$_2$O). Less O$_2$ consumed by the decomposition of organic matter means more O$_2$ left in the atmosphere.

** Pelagic organisms live in the water column. Benthic organisms live on or near the seafloor.
The earliest multicellular fossils of land plants are found in rocks of Ordovician age approximately 475 million years old (Fig. 9A). The advantage of a terrestrial mode of life for the earliest land plants was threefold: (1) abundant sunlight, (2) abundant nutrients, and (3) absence of predators. The arthropods (crabs, spiders, scorpions, centipedes...) were the first animals to colonize the land. The earliest fossil evidence is based on trackways preserved in Late Cambrian-Early Ordovician eolian sandstone in Canada (Fig. 9B). Optimal environmental conditions and the lack of predators enabled some arthropods to reach a gigantic size during the Carboniferous (e.g. 70-cm large dragonflies, 2-m long millipedes). The first terrestrial vertebrates (tetrapods) evolved from a group of Silurian fishes. A good example of an early tetrapod is *Ichtyostega* which lived during the Devonian period (Fig. 9C). Reptiles appeared during the Carboniferous. In this group, the dinosaurs became extremely diverse and some grew larger than any other land vertebrates during the Jurassic and Cretaceous. They flourished for a hundred million years before vanishing 65 million years ago (see section 12.6.2.). Their only living descendants are the birds. The last 65 million years is known as the Cenozoic and is marked by the diversification of mammals. This group includes the genus *homo* (us and our ancestors), the earliest representative of which appeared around 2.3 million years ago.

![Figure 9: 475 million years old spores of an early land plant (A), Late Cambrian or Early Ordovician trackways of arthropods, and a 3D image of a skeleton of *Ichtyostega* (sources: A. Wellman et al., 2003, B. McNaughton et al., 2002, C. Pierce et al., 2012)](image)

### 12.6. Mass extinctions

The history of life is characterized by five major extinction events (Fig. 10). Each were followed by a period of diversification. Although the biosphere has been hit by 5 major mass extinction events, the global species richness has increased through time. We will now focus on two of these mass extinction events: (1) the most severe of all: the end-Permian mass extinction; and (2) the most famous of all: the end-Cretaceous mass extinction.

#### 12.6.1. End-Permian mass extinction

The end-Permian mass extinction occurred 250 million years ago. Roughly 90% of all species went extinct. The crisis affected both marine and terrestrial ecosystems. The Permian is a geologic period which began 300 million years ago and ended 50 million years later. The geography of the Permian is characterized by the presence of a single supercontinent called Pangaea. The reduction of shallow marine environments enhanced the competition for resources and may have contributed to species extinction. However, other factors must have played a role to cause a mass extinction of such a magnitude. One of the most likely factors is a catastrophic volcanic eruption which disrupted global climate. Such an intense and long-lasting eruption has never been witnessed in human history. However, the geological record shows
evidence that gigantic eruptions have occurred several times in the past. These eruptions have produced tremendous amounts of basalt called **flood basalt**. Flood basalts cover extensive regions of the crust called flood basalt provinces. One such province, the largest one on a continent, is found in Siberia and its age coincides with the end-Permian extinction: the Siberian flood basalts (or **Siberian Traps**) which have accumulated during a million years and covers approximately 2 million km$^2$.

What would be the impact of a catastrophic volcanic eruption on global climate?

1. On the short term: massive amounts of volcanic ashes released in the upper atmosphere would reflect some of the incoming sunlight and cause a short-lived global cooling.

2. Acid rains would result from the reaction between volcanic gases and condensing water droplets ($\text{H}_2\text{SO}_4$, $\text{HCl}$, $\text{HF}$).

3. On the long term: the release of massive amounts of CO$_2$ in the atmosphere would trigger a global warming. The consequences of global warming are complex. For example, global warming may disrupt the ocean circulation and prevent the renewal of O$_2$ in the deep ocean*. The warming of seawater also decreases the solubility of O$_2$ in the oceans. Moreover, the stability of **methane hydrate** present in deep sea sediments and in high-latitude permafrost may be affected by global warming. Methane hydrate is composed of water ice with molecules of methane trapped in the crystal lattice of the ice. Methane is mainly produced by methanogen archaea through anaerobic respiration involving the byproducts of the decomposition of organic matter. During global warming, there is risk of destabilization of methane hydrate which may lead to a massive input of methane in the atmosphere. Methane is a very potent greenhouse gas which would further enhance global warming and amplify the destabilization of methane hydrate, creating a **positive feedback loop**.

12.6.2. End-Cretaceous mass extinction

The end-Cretaceous mass extinction 65 million years ago is the most famous mass extinction event because it caused the extinction of dinosaurs. Unlike the other mass extinction events, the cause of the end-Cretaceous mass extinction is relatively well constrained. There are several convincing lines of evidence supporting the hypothesis of a large **meteorite impact** as the main cause.

The Cretaceous-Tertiary boundary (known as the K-T or K-Pg boundary) is characterized by a distinctive layer that can be followed worldwide and which contains a high concentration of Iridium (rare in earth’s crust but abundant in meteorites) and components indicative of a meteoritic impact.

---

* One of the main source of deep ocean water is in the North Atlantic where cold and salty (dense) water sinks to the bottom of the ocean and begins its journey to the south hemisphere (= thermohaline circulation). If the ocean surface warms, the downwelling of O$_2$-rich surface water in the North Atlantic and elsewhere may be disrupted and prevent the oxygenation of the deep ocean.
(shocked quartz, tektite, nanodiamonds). The crater associated with the K-T boundary has been located near the Yucatan Peninsula in Mexico (Chicxulub Crater). The meteorite that produced the Chicxulub Crater must have been at least 10 km in diameter. The consequences of such a large meteorite impact on the environment are multiple and take place at various time and spatial scales (table 1).

Figure 10: Evolution of biological diversity over the past 600 million years (source: Understanding Earth 6th edition)

Table 1: Environmental effects of the meteorite impact at the K-T boundary (source: Toon et al., 1997)