IOP Publishing

EPS Grand Challenges
Physics for Society in the Horizon 2050
Mairi Sakellariadou, Claudia-Elisabeth Wulz, Kees van Der Beek, Felix Ritort, Bart van Tiggelen, Ralph Assmann, Giulio Cerullo, Luisa Cifarelli, Carlos Hidalgo, Felicia Barbato, Christian Beck, Christophe Rossel and Luc van Dyck

Chapter 4

Physics for understanding life

Patricia Bassereau, Angelo Cangelosi, Jitka Čejková, Carlos Gershenson, Raymond Goldstein, Zita Martins, Sarah Matthews, Felix Ritort, Sara Seager, Bart Van Tiggelen, David Vernon and Frances Westall

4.1 Introduction

Felix Ritort[1] and Bart Van Tiggelen[2]
[1]University of Barcelona, Barcelona, Spain
[2]Laboratoire de Physique et de Modélisation des Milieux Condensés, University Grenoble Alpes/CNRS, Grenoble, France

Infinitely many challenges exist for physics and science in general to understand life and human behaviour. Many of them were identified already a long time ago: What is the nature of the human mind? Does life exist elsewhere in the Universe? How did life emerge on Earth? Did molecules essential for life come out of space? The huge progress in physics, chemistry, biology, computer science, and astrophysics over the last decennia has made it clear that answers to these complex questions have become within reach. They will emerge from the next generation of scientists, forming collaborations using advanced technologies where frontiers between disciplines will eventually disappear.

We currently have a lack of knowledge of the early environmental conditions on Earth when life emerged, especially because no well-preserved crust is known to exist older than 3.5 billion years. Hopefully, a remnant of the ancient crust may be discovered under the Antarctic ice cover. Removing the uncertainties around the forcing of climate by the young Sun is a second major challenge. The Sun–Earth connection must have played an essential role in the prebiotic chemistry, before life was created. Finally, we are the first generation able to truly search for extraterrestrial life. Today we can send landers to targets in the Solar System and remotely probe the atmospheres of exoplanets for biosignatures. Many new fly and probe missions are under study or even under design. Understanding how life is created on Earth and unveiling life elsewhere is a tremendous challenge for mankind that will change drastically our vision of the Universe.

In the past decades, physics has faced the challenge of unravelling hidden principles that explain the marvelous complexity of living matter. While biology and biochemistry have made huge steps over the past two centuries, a quantitative understanding of life itself remains a formidable challenge. Life is permanently out-of-equilibrium, and despite the major progress in physical and chemical sciences, we still do not understand how to produce life from raw materials. Is it reasonable to expect the advent of new physical laws or principles emerging from biological studies? Is biology going to unveil a deep nexus with physics one day, as chemistry already did a century ago? Finally, does natural evolution have the flavor of physical law, or is it just conditional for life to exist? Almost sure, all these questions will be at the core of physics research in this century.

Artificial intelligence and artificial life aim at answering to these questions posing new and provocative challenges for the so-called fourth industrial revolution. Can we build machines that solve tasks like humans do? Can we empower machines and robots with brain-inspired algorithms? Will we ever understand how intelligence and consciousness work and how we deal with uncertainty? Besides designing machines and algorithms that behave as living beings, we would like to implement living matter in the lab. Physics offers uncountable possibilities to build artificial wet life. Compared to living cells, synthetic cells exhibit limited properties and functionalities, yet there may be ways to organize living matter beyond what we see in nature. A remarkable feat would be to develop artificial autonomous systems with open-ended evolution, the feature that makes animate matter so different from inanimate matter.

This chapter presents a collection of articles on these four topics by exposing the main challenges that lie ahead in these exciting fields.

4.2 Searching for life in the Universe: what is our place in the Universe?

Sara Seager[1] and Zita Martins[2]
[1]Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA
[2] Centro de Química Estrutural, Institute of Molecular Sciences and Department of Chemical Engineering, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal

For thousands of years, inspired by the star-filled dark night sky, people have wondered what lies beyond Earth. Today, the search for signs of life is a key factor in modern-day planetary exploration, both for *in situ* exploration of our own Solar System's planets and moons, and by remote sensing via telescopes of exoplanets orbiting nearby stars. Planetary *in situ* measurements enable us to search directly for organics and even life itself. The holy grail of detection of life in our Solar System would be to detect a 'fingerprint' of present or/past extraterrestrial life, either in the atmosphere or regolith of planets, or in the oceans of the icy moons of Jupiter and Saturn. Beyond our Solar System, we aim to detect a gas in an exoplanetary atmosphere that might be attributed to life. A suitable 'biosignature gas' must be able to accumulate in an atmosphere against atmospheric radicals and other sinks, have strong atmospheric spectral features, and have limited abiological false positives. This review summarizes the organic compounds that are representative of potential biosignatures in planets and moons, and how to distinguish these from prebiotic chemistry, it recalls the growing list of potential biosignature gases, as well as the next generation of telescopes under construction or in development that have the capability to detect biosignature gases in exoplanet atmospheres.

4.2.1 General overview

4.2.1.1 Introduction and motivation through the lens of time

As a species we have wondered about life beyond Earth for millennia, since at least the time of the Greek philosophers (e.g., [1]). Hundreds of years ago some accepted the concept of life elsewhere. A few of the Founding Fathers of the United States of America are quoted with public statements in the late 1700s saying that stars were suns with planetary systems[1] [1]. After Bessel had invented new tools for astrometry and quantified distances to stars, US school children in the mid-1800s Midwest were reportedly taught not only about those stellar distances but also that there are humans out there in planetary systems surrounding those stars [2].

Over a century ago, astronomers laid the foundations of the astronomical search for life by remote sensing. Arcichovsky [3] proposed looking for vegetation

[1] '*The probability, therefore, is that each of these fixed stars is also a Sun, round which another system of worlds or planets, though too remote for us to discover, performs its revolutions, as our system of worlds does round our central Sun.*' Thomas Paine, in [1].

signatures in the Moon's Earthshine as a reference case for vegetation or chlorophyll searches on other planets. In 1930, the famous astronomer Jeans described the concept of oxygen as an atmospheric biosignature gas[2].

The search for signs of life on Mars also began over one hundred years ago with the sensational reports of canals on Mars [5, 6]. About half a century later, the search for signs of life by way of remote sensing spectroscopic analysis yielded a report of the detection of the 3.4 micron absorption of vegetation on Mars' surface [7, 8], which was later found to be deuterated water (HDO) in the Earth's atmosphere [9]. In the 1960s and 1970s, the first spacecrafts, the Mariners, took images of Mars on orbit, revealing several extinct craters and volcanoes, but no canals or vegetation [10]. Hence, ideas about canals and vegetation on Mars were completely dropped.

The *in situ* search for life beyond Earth began in earnest with the Mars Viking lander missions [11]. No organic compounds were detected on the Martian regolith above a threshold level [11]. This was due to two reasons: (i) technological problems, as the gas chromatography–mass spectrometer (GC–MS) onboard the spacecraft was not able to analyze and detect organic molecules in the regolith [12], and (ii) the location of the search—on the surface of Mars—where several reactions (that were unknown at the time) destroy organic compounds. These destructive reactions may happen because of UV radiation [13–22], cosmic rays [23, 24] and oxidation species [25], such as perchlorates. Indeed, perchlorates were detected directly on the Mars regolith by the Wet Chemistry Laboratory (WCL) of the Phoenix spacecraft, and indirectly by the Thermal Evolved Gas Analyser, with the analysis indicating the thermal decomposition of perchlorate [26, 27].

While today sample return and *in situ* search for ancient and present-day life has become within reach for Solar System bodies (including Mars, Europa, Enceladus, and Titan), the search for signs of life on exoplanets will remain in the domain of remote sensing. We know of thousands of planets beyond our Solar System, orbiting stars other than the Sun, called exoplanets (figure 4.1; [28]). Nearly every star is expected to have planets of some kind. While Solar System copies are hard to find due to observation selection bias, planetary systems like our own appear to be rare. Figure 4.1 shows that the observation of an Earth analog (an Earth-size planet in an Earth-like orbit around a sun-like star) remains technologically out of reach. Hence, today's search for habitable worlds and signs of life is focused on a planet category easier to discover and to observe than an Earth analog, such as Earth-sized or larger planets orbiting small red dwarf stars. A planet orbiting a red dwarf has a larger signal with currently favoured detection techniques than a planet of the same size orbiting a sun-sized star or even larger. Nonetheless, committees dating back to the 1960s discussed space mission designs to search for Earth analogs. NASA's

[2] '*It seems at first somewhat surprising that oxygen figures so largely in the Earth's atmosphere, in view of its readiness to enter into chemical combination with other substances. We know, however, that vegetation is continually discharging oxygen into the atmosphere, and it has often been suggested that the oxygen of the Earth's atmosphere may be mainly or entirely of vegetable origin. If so, the presence or absence of oxygen in the atmosphere of other planets should shew whether vegetation similar to that we have on earth exists on those planets or not.*' [4].

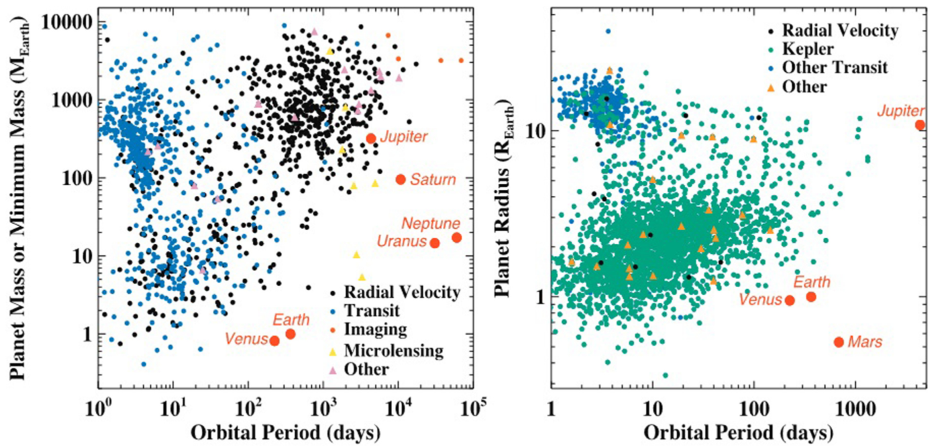

Figure 4.1. The distributions of known exoplanets as a function of their mass and period (left) and radius and period (right). The observed semi-major axis is converted to period via Kepler's third law. The color coding denotes the method by which planets were detected. Some of the features in these diagrams are real, but many are due to the current observational selection effects. Notably, the ~4300 confirmed exoplanets have an astonishing diversity in mass, size, and orbital period. Earth analogs are isolated on this log–log diagram, supporting the point that they are technologically just out of reach. Figure reproduced from [29].

Terrestrial Planet Finder studies[3] in the early 2000s led to the space-based, high-contrast, direct imaging mission concepts of today that will be able to find and identify a true Earth analog.

The study of exoplanetary atmospheres relies on remote sensing, blending decades of solar-system-planet foundational work with new techniques developed to cope with the exoplanet atmospheres' low spectral resolution and the lack of spatial or vertical resolution. With our first generation capability of observing atmospheres of rocky exoplanets of suitable temperatures for life, eventually, the goal is to search for 'biosignature gases', i.e., gases produced by life that can accumulate in exoplanetary atmospheres to remotely detectable levels. The main assumption we make is that life elsewhere uses a chemistry as life on Earth does, to extract, store, and release energy for metabolism, and in this process generates waste products as biosignature gases. Astronomical spectroscopy begun with the detection of dark lines in the solar spectrum by Baker in 1801 and continues today with stellar, planetary, and exoplanetary spectroscopy of ever-increasing sophistication. The search for signs of extraterrestrial life will continue for the next century and beyond.

4.2.1.2 Origin of life, habitability, and signs of life

Since antiquity, the question of the origin of life has been under discussion. In the beginning of the 20th century, Oparin and Haldane [30–32] suggested that the atmosphere of the early Earth was strongly reduced, mainly composed of methane,

[3] http://science.jpl.nasa.gov/projects/TPF/

ammonia, water vapor, and molecular hydrogen [33]. According to Oparin and Haldane, spark discharges in this atmosphere would generate organic compounds, which would later accumulate in the ocean of the primitive Earth. Amino acids were then experimentally synthesized in the laboratory by Stanley Miller using equipment that mimicked the atmospheric and ocean conditions of the primitive Earth suggested by Oparin and Haldane [34, 35]. However, towards the end of the century, new data (e.g., geology information, composition of volcanic gases, models of accretion of the Earth) showed that the atmosphere of the primitive Earth was not reduced, but was composed of nitrogen, carbon dioxide, and water vapor [36–40], so that this could not be the mechanism to produce significant amounts of organic molecules [41–45]. Presently, the exact time and conditions associated with the origin of life on Earth are the subject of intensive debates in the scientific community (e.g., [46–57], section 4.6). Despite not knowing the exact mechanism of formation, the first cell—the basic unit of life—must have needed water and organic molecules, either formed abiotically on Earth or exogenously delivered [58]. The origin of life was undoubtedly not unique to our planet, and it makes sense to search for habitability conditions in other worlds.

The search for habitable worlds in our Solar System (figure 4.2) focuses on planets and moons that have three main requirements: The presence of (1) a source of energy, (2) elements essential for life that support a geochemical mechanism, and (3) liquid water [59]. Liquid water is a solvent that is able to solubilize ions (cations and anions), which is fundamental to generate ionic gradients between the inside and outside the cells. In addition, water is also expected to be the most abundant liquid in the cosmos [60]. In our Solar System, Mars is considered to have the right habitability conditions as it was much wetter in its past, during the Noachian and early Hesperian Eons [61, 62]. Furthermore, icy moons of our Solar System are modelled to have rocky cores [63, 64], whose interaction with the subsurface ocean is a fundamental condition to generate geochemical reactions. Europa and Enceladus

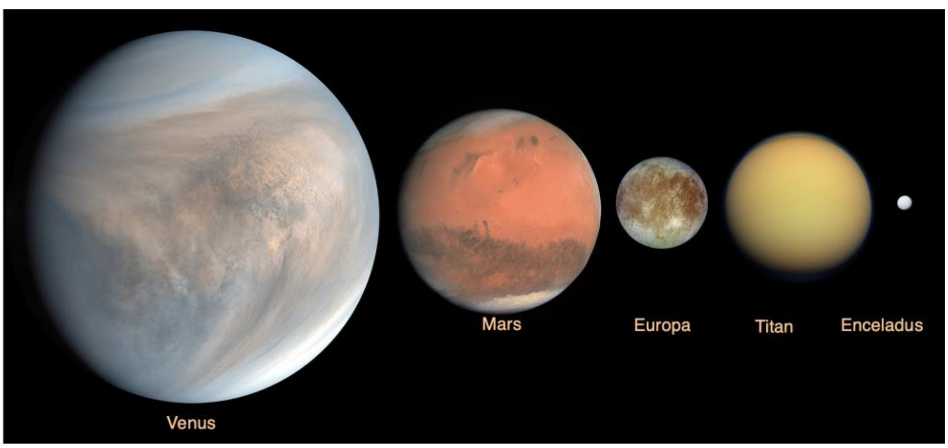

Figure 4.2. Potentially habitable planets or moons in our Solar System. Not to scale. Planet images credit, Venus: JAXA, others: NASA.

are amongst the icy moons in which water/rocky core interactions are present [65, 66], thus increasing the potential for finding signatures of life at these locations.

We may, on the whole, exclude giant exoplanets from the list of life-supporting planet types as there is no solid surface below their atmospheres but rather a compressed phase of hydrogen and helium at temperatures far too high to support covalent bonds. Water vapor in a small rocky exoplanet atmosphere still is a key gas in our search for habitable worlds.

Organic matter is widespread throughout the Universe, so to successfully detect extraterrestrial life and its biosignatures, it is necessary to first distinguish between abiotic matter and signatures of past or present life. Abiotic matter that may have played a role in prebiotic chemistry just before life originated in our Solar System may be studied by laboratory analysis of organic residues produced by simulated photo- and thermo-processing of icy mixtures. They are considered as analogues for the organic material synthesized in interstellar or circumstellar icy grains [67–72], laboratory analysis of meteorites, micrometeorites, interplanetary dust particles (IDPs) and Ultra-Carbonaceous Antarctic Micrometeorites (UCAMMs) [73–85], on-site analysis of small planetary bodies (comets and asteroids), and laboratory analysis of sample return missions from these small planetary bodies (see [86, 87] and references therein).

A biosignature is a substance and/or a chemical pattern whose origin requires a biological agent [88, 89]. A cell is made up of i) proteins, which in turn are made up of amino acids, ii) genetic material, which is made up of nucleotides which in turn are made up of nucleobases, ribose, and phosphates, and iii) a membrane consisting of amphiphilic molecules spontaneously self-assembled into vesicles in water. Proteins, genetic material, i.e., RNA, and DNA would be ideal biosignatures, but they easily degrade under oxidizing and radiation conditions [90]. Furthermore, and with very few exceptions, life on Earth uses L-amino acids and D-sugars, while on meteorites mixtures of D- and L-amino acids have been detected (for a review [82, 83]). So chirality may be used as a tool to distinguish between abiotic organic molecules and products of past or present life. In addition, stable carbon, nitrogen, hydrogen, and sulfur isotope data of individual organic molecules may also be used to make this distinction. Abiotic molecules formed in extraterrestrial environments have been found to be enriched towards the heavy isotopes such as deuterium, ^{13}carbon, and ^{15}N [82, 83]). A likely biosignature deriving from molecular attributes of amino acids and sugars would require to exhibit chiral asymmetry, as well as a simple molecular distribution with structural isomer preference, and a light isotopic composition [91]. A summary of targeted biosignatures and their level of priority for future space missions in our Solar System is provided by Parnell *et al* [92].

To search for signs of life in an exoplanetary atmosphere, we make the assumption that like on Earth, life elsewhere will use chemistry to extract energy from the environment, to store energy, and will use the energy in a way that might generate a by-product gas. In other words, we assume that life elsewhere produces by-product 'biosignature gases'. The biosignature gases must accumulate in an exoplanet atmosphere to levels that astronomers can detect using next-generation ground- and space-based telescopes. The general consensus is to look out for gases that are present in too high amounts for them to be produced by abiotic chemistry, and that are also incompatible with their atmospheric and planetary environment. A touted example is that the simultaneous presence of CH_4 and O_2 could be a clear indicator of life, because CH_4 (a reduced gas) would be rapidly destroyed by reactions with the types of molecules present in a highly oxidized environment. However, both were never seen to be simultaneously present at high enough levels on Earth for this idea to be validated [93, 94]. Researchers are working hard to come up with many scenarios where a particular gas could be attributed to life given the context of a certain planetary environment. Part of this flourishing research area is complicated by false positives. This happens, for instance, when a biosignature gas is in reality produced by abiotic processes. For a comprehensive review see [95]. A key challenge is to overcome the problem that many gases of real interest may not be detectable and that those that can be detected may not be robustly associated with life without excluding false-positive scenarios.

There is a growing list of gases that are being considered as candidate biosignatures. These gases are input in computer programs to simulate their survival against destructive photochemistry and to assess their detectability by remote observations with future telescopes. The gases include dimethyl sulfide (DMS or $(CH_3)_2S$) [96–98]; nitrous oxide (N_2O) [99]; methyl chloride (CH_3Cl) [100, 101]; phosphine (PH_3) [102]; isoprene (C_5H_8) [103]; and ammonia NH_3 [104, 105] and other amines [104]. The unique approach by Seager *et al* (2016) aims to exhaustively sort through all classes of potential biosignature gases in order to find new biosignatures. More details can be found in reviews [95, 106, 107].

4.2.1.3 Solar System planets and moons: current status, and future plans

The planet Venus seems to be an unlikely abode for life due to the massive CO_2 greenhouse atmosphere creating surface temperatures of 700 K, too hot for complex molecules to survive, and thus too hot for life of any kind. However, for over half a century [108], people have speculated that life may exist in an aerial biosphere, populating the Venus atmosphere's clouds at altitudes of 48–60 km above the surface where temperatures are suitable for life. The clouds are permanent and cover Venus completely. Yet, the Venus cloud environment is undeniably harsh (for a review see [109]), both very dry and very acidic. The clouds are composed of liquid sulfuric acid which are more than ten orders of magnitude more acidic than the most acidic environment on Earth [110, 111]. A number of intriguing atmospheric anomalies on Venus have persisted for decades and may or may not be connected to life. Some are robust, including the unidentified but very strong ultraviolet absorber (e.g., [112]); the *in situ* detection of O_2 by two different probes, Pioneer 13 [113], and Venera 14 [114]; and the anomalously low abundance of H_2O and SO_2 [115]. Other anomalies are still tentative, such as the controversial detection of PH_3 from ground-based radio telescopes [116–118] that is also found in re-analyzed data from the Pioneer 13 probe [119], and the suggested detection of NH_3 (Pioneer 13: Mogul *et al* 2021, and Venera 8: [120]). It is interesting to note that the last entry probes were done nearly four decades ago: the Pioneer Venus probes in 1979 (e.g., [121]) and the Russian VEGA balloons in 1985 (e.g., [122]). Currently, the Japanese

Aerospace Exploration Agency (JAXA) Venus Akatsuki is orbiting Venus, and the European Space Agency (ESA) Venus Express orbited for nearly a decade until 2015. Both have returned valuable data on Venus' atmospheric dynamics and composition as well as on the cloud composition and structure. A number of missions are either being proposed, are under study, or are planned by a range of countries including the USA, Russia, Europe, Japan, and India.

The surface of Mars contains oxidized species such as perchlorates and chlorates (e.g., chlorinated hydrocarbons), which were detected by the Sample Analysis at Mars (SAM) of the Mars Science Laboratory's (MSL) in the Gale Crater and Sheepbed mudstone at Yellowknife Bay [123–127]. These molecules destroy any potential biosignature, and therefore, it is not likely to detect life on the surface of Mars, even if it was there. In February 2021, the Mars2020 mission—which will investigate the past habitability conditions, the potential for past life on Mars, and the preservation of biosignatures—landed the rover Perseverance and its helicopter Ingenuity on Mars. The mission will store samples that will later be returned to Earth on a future Mars sample return mission. ESA and the Russian space agency Roscosmos have launched the ExoMars Trace Gas Orbiter (TGO) into the orbit of Mars. ESA will launch the ExoMars rover—named Rosalind Franklin—to search for biosignatures at various depths of Mars up to 2m, but it is unlikely that this will happen before 2028 [128]. Since April 2018 the TGO has been analyzing the atmosphere of Mars to establish the presence of methane and other trace gases [129]. The previous numerous reports of CH_4 come from various instruments, including three different large ground-based telescopes (e.g., [130]), the Prime Focus Spectrograph (PFS) instrument on the Mars Orbital Express (e.g., [131, 132]), the thermal emission spectrometer (TES) onboard the Mars Global Surveyor (e.g., [133, 134]), and from the Tunable Laser Spectrometer (TLS) component of the Sample Analysis at Mars (SAM) instrument on the Mars Curiosity Rover [135, 136]. However, observations from the TGO did not find any methane in any hemispheres beyond 0.05 parts per billion by volume [137]. The presence of CH_4 can only be reconciled with the non-detection from the TGO if concentrated amounts produced locally at the near-surface are destroyed by an unknown process that can rapidly remove or sequester methane from the lower atmosphere before it spreads globally [137]. If methane is indeed present, it is not known (see figure 4.3) if the source is geological in origin, due to early Mars serpentinization reactions leading to the formation of hydrogen and methane, or possibly even biological, for instance produced by methane-producing microorganisms.

The favorable habitability conditions of the Jovian and Saturnian icy moons increase the probability for these moons to accomodate life. The Saturnian moon Enceladus contains a silicate core and an ocean beneath its surface [138, 139]. Indirect analysis of this ocean by the Cassini space mission showed the existence of a mixture of small (below 50 atomic mass units) and complex macromolecular organic compounds (beyond 200 atomic mass units) [140, 141]. The European Space Agency is planning to launch in the next decade the Jupiter Icy Moon Explorer (JUICE) to explore the Jovian system—Ganymede, Callisto, and Europa—and its organic content related to biosignatures in Europa [142].

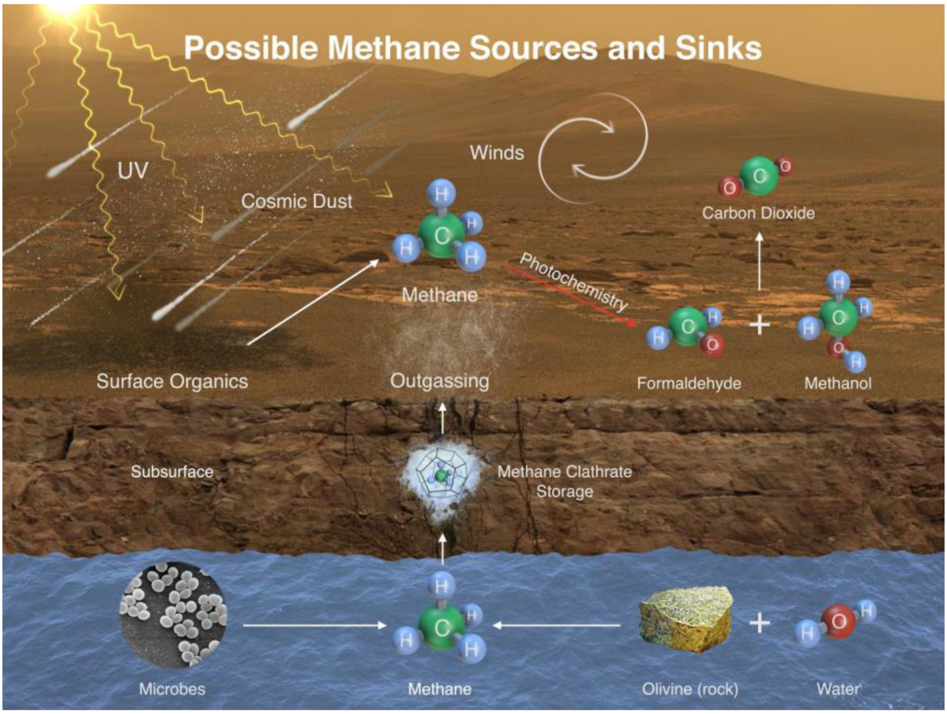

Figure 4.3. Possible sources and sinks of Mars' atmospheric methane (biological and/or geological). Credit NASA/JPL-Caltech/SAM-GSFC/University of Michigan. https://www.jpl.nasa.gov/images/possible-methane-sources-and-sinks.

4.2.1.4 Building blocks of life on small Solar System bodies

The building blocks of life have been analyzed and detected in several small Solar System bodies, such as comets and asteroids. The comet Halley contains an organic mantle composed of highly unsaturated compounds [143]. In the beginning of the 21st century the Stardust mission collected particles of comet 81P/Wild 2, which were returned to Earth. Analysis showed deuterium and ^{15}N excesses in some of the comet particles, with both aromatic and non-aromatic compounds [144, 145], including refractory material with highly aromatic organic matter [146]. Glycine (with a carbon isotopic composition of 29‰ ± 6‰), methylamine, and ethylamine were detected in the particles of comet 81P/Wild 2 [145, 147]. The dust particles of comet 67P/Churyumov–Gerasimenko were analyzed *in situ* by the Rosetta space mission and showed solid organic matter similar to the insoluble organic matter (IOM) of carbonaceous chondrites [148]. The nucleus of the comet 67P/Churyumov–Gerasimenko contained non-volatile organic macromolecular materials [149], carbon-rich species (e.g., alcohols, carbonyls, amines, nitriles, amides, isocyanates, the polymer polyoxymethylene [150, 151], phosphorus, and glycine) [152].

Samples were collected from asteroids and returned to Earth to be analyzed. The 25143 Itokawa near-Earth asteroid was explored by the Hayabusa spacecraft from the Japanese Aerospace Exploration Agency (JAXA) [153], and some of its particles were found to be carbonaceous, also containing nitrogen, oxygen, and trace amounts of fluorine and sulfur [154]. More recently, the Hayabusa2 mission touched down on the near-Earth C-type asteroid 162173 Ryugu, and successfully returned samples to Earth in December 2020. A variety of molecules were detected and identified [155–157]. In 2023 scientists received samples from another asteroid, the B-type asteroid 101955 Bennu, which was visited by the OSIRIS-Rex space mission [158]. Both asteroids and comets have delivered significant amounts of extraterrestrial organic molecules to Earth, which may have contributed to the feedstock of organic matter necessary for the first living organisms to be created on our planet.

4.2.1.5 Exoplanets: current status, and future plans

The current status of the search for signs of exoplanetary life is an exciting one. We are on the brink of the launch of a brand-new space telescope, one that has been many decades in the making. The James Webb Space Telescope (JWST; [159]) was launched in December 2021. JWST is an internationally supported mission, primarily NASA, with contributions from the ESA and the Canadian Space Agency (CSA). JWST is an infrared telescope that has more than seven times the collecting area of the Hubble Space Telescope and is orbiting more than 1 million miles from Earth at the L2 Earth–Sun gravitational balance point, an ideal vantage point for observational astronomy because it is away from the contaminating light and heat of Earth.

Because JWST is a general-purpose observatory, it will be used by hundreds of astronomers to study transiting exoplanets and their atmospheres. Transiting planets are a set of exoplanets with fortuitously aligned orbits such that the planets pass in front of their stars as seen from our viewpoint (figure 4.4). During transit, the starlight shines through the planet’s atmosphere. Some wavelengths of light make it through

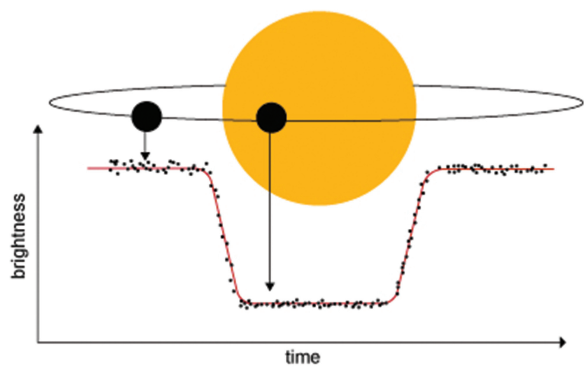

Figure 4.4. Transit planet schematic. Transiting planets are those that orbit in front of the star as seen from our viewpoint and cause a tiny drop in the star’s brightness of a very characteristic shape and duration (graph). Note that stars other than the Sun are not spatially resolved; the cartoon star and planet are for illustration only. We know of thousands of transiting exoplanets. Astronomers are able to study the atmospheres of transiting planets because during transit the exoplanets atmosphere absorbs some wavelength-specific starlight, depending on the gases present in the exoplanet atmosphere (not shown). Figure credit: https://www.open.edu/openlearn/ocw/mod/oucontent/view.php?id=67466&section=3.1.

while others are absorbed by gases—enabling us to identify gases in an exoplanet atmosphere [160]. Today, we use the Hubble Space Telescope to study mostly hot, giant exoplanets. JWST holds promise to push observations down to smaller rocky exoplanet atmospheres—the kind of planets that might host life. JWST is limited to transmission spectra of planets transiting red dwarf stars—a very common type of star much cooler and smaller than the Sun. We should keep in mind that JWST was not designed for exoplanet science, and that we do not yet know its systematic noise floor. Nonetheless, JWST is already programmed to observe the so far most favorable system of potentially habitable planets that we can observe, those orbiting the ultracool red dwarf star TRAPPIST-1 [161] provided that the planets have maintained their atmospheres against the high-energy radiation coming from the host star [162].

Beyond JWST are planned the very large next-generation ground-based telescopes now under construction. These include the Extremely Large Telescope (ELT, with 39 m aperture diameter) [163], the Thirty Meter Telescope (TMT, 30 m aperture diameter) [164], and the Giant Magellan Telescope (GMT, 20 m aperture diameter) [165, 166]. These ELTs will be able to study atmospheres of exoplanets orbiting small red dwarf stars— though not limited to transiting exoplanets. The telescopes will operate at near-infrared wavelengths and must have special instruments to block out the starlight so we can see the planet directly, a method called direct imaging. Direct imaging requires special control of the telescope optics, called adaptive optics, to counteract Earth's atmospheric turbulence that causes distortions to the images. The challenge for directly imaging exoplanets in the habitable zones of M dwarf stars is to overcome the huge brightness difference between the planet and star: the light reflected by the planet is 10^7–10^8 times fainter than the star. A very suitable planet for the ELT observations is the Earth-mass planet orbiting the star nearest to the Sun, Proxima Centauri b [167].

The lofty goal to discover an Earth analogue exoplanet is extremely challenging. By Earth analogues we mean Earth-sized planets in Earth-like orbits around Sun-like stars. The challenge is that an Earth analogue in reflected light is ~10 billion times fainter than the host star (see figure 4.5). We must be able to suppress the starlight by such a tremendous factor. Only in space, above the blurring effects of Earth's atmosphere can telescopes meet this challenge for a significant number of stars. Moreover, for space telescopes, the search for biosignature gases will not be contaminated by the same gases already present in the Earth's atmosphere.

One promising and fancy concept to suppress the starlight is Starshade, a giant specially shaped screen tens of meters in diameter that would fly in formation thousands of kilometeres in front of a space-based telescope (figure 4.6). Starshade would block out the starlight so that only planet light enters the telescope. Starshade was first conceived in the 1960s [168] and has been revisited every decade since then until the mid-2010s, when a serious plan to make Starshade a reality was formulated by a NASA Study team [169]. Today there are several Starshade mission concepts of different sizes, including the Starshade Rendezvous Mission and the Habitable Exoplanet Observatory ([28]; figure 4.7) and the projects are awaiting funding. A related concept is the internal coronagraph, a starlight suppression device inside the telescope, such as described for the Large Ultraviolet Optical

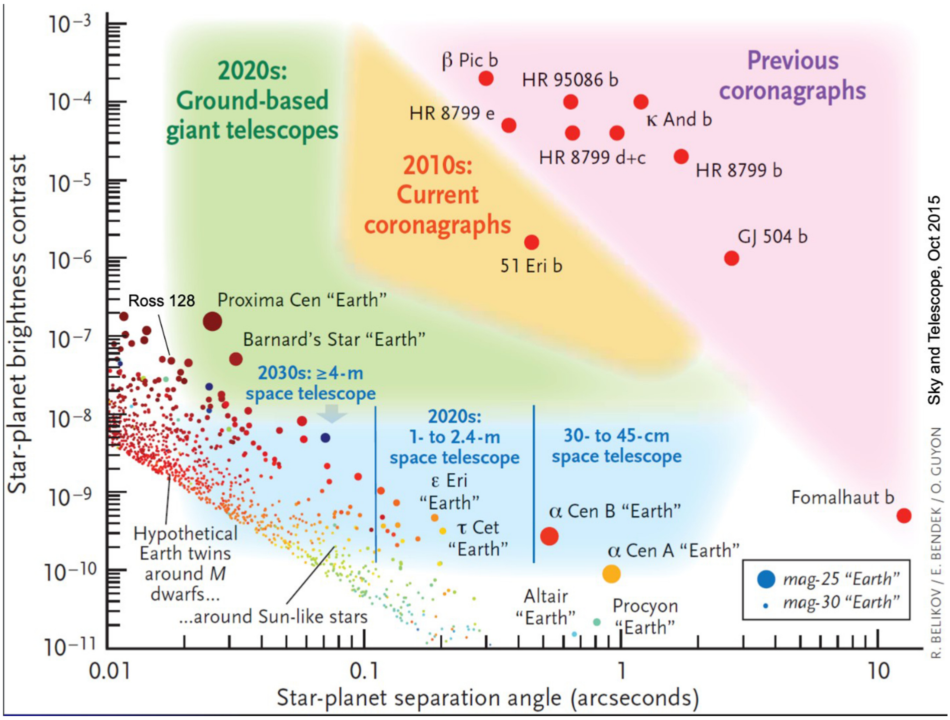

Figure 4.5. Overview of the field of exoplanet 'direct imaging' in visible and infrared light. The *x*-axis is the planet-star projected separation on the sky in angular units and the *y*-axis is the ratio of the planet's reflected light to the starlight. The point is that any telescope has to be able to suppress starlight to very significant levels for a planetary atmosphere to be detected by direct imaging. Capabilities of past, current, and future direct imaging instruments are shown. Red dots in the upper right are known as directly imaged giant exoplanets. Figure credit: Belikov, Bendek and Guyon, Sky and telescope 2015.

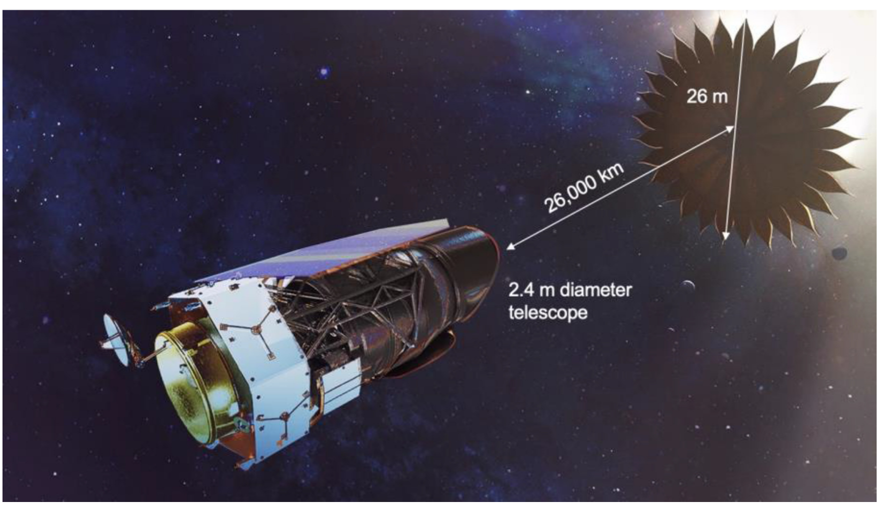

Figure 4.6. The Starshade Rendezvous Probe Mission consists of the Roman Telescope and a formation-flying starshade. The mission will enable direct imaging of the habitable zones of nearby Sun-like stars with sufficient sensitivity to detect and identify Earth-like exoplanets. Image not to scale. Credit: NASA.

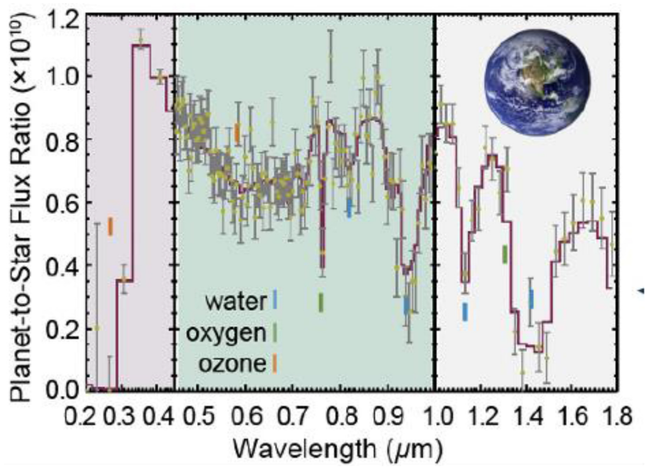

Figure 4.7. Simulated Earth-like planet spectra with the HabEx Observatory. The majority of HabEx-discovered planets will be well-characterized, with relatively high signal-to-noise ratio (SNR $\geqslant$ 10), and for a spectral resolution $R\,\lambda/\Delta\lambda$ of about 7 in the UV channel from 0.20 to 0.45 μm, $R \sim 140$ spectra in the visible channel from 0.45 to 1.0 μm, and $R \sim 40$ spectra in the IR channel from 1.0 to 1.8 μm. Figure from the HabEx Final Report [28].

Infrared Surveyor [170]. There is one hybrid Starshade coronagraph mission concept, HabEx. The bright, nearby Sun-like star tau Ceti is reported to host a multiple planet system [171, 172], including two habitable planets suitable for space-based high-contrast direct imaging.

4.2.2 Challenges and opportunities for 2050

4.2.2.1 Solar System planets and moons: 2050 and beyond

Currently there is an ongoing discussion on how best to obtain pristine samples from Solar System bodies where life may have started, either by future on-site missions or by sample-return missions. The decision will be based on time, cost, risk of contamination, and payload [87]. While on-site analysis will constrain the risk of contamination, the equipment onboard will be much more limited than what can be done in laboratories on Earth. On the other hand, sample-return missions will allow the use of a diversity of state-of-the-art equipment, but the risk of contamination of the samples is high, and curation facilities need to be in place. Even if contamination controls are implemented, minimal levels of terrestrial contamination will always occur, so that scientists must be able to distinguish between extraterrestrial organic matter and terrestrial organic matter [173]. There are several targets for 2050 and beyond, that will be briefly described below.

There is a tremendous opportunity to visit Venus with instruments with today's technology to search for signs of life or even life itself directly in the Venus atmosphere (venuscloudlife.com). Even though about three dozen space missions have already gone to Venus, there have never been any direct probes of the atmosphere since the NASA Pioneer and the Russian Venera Probes have flown

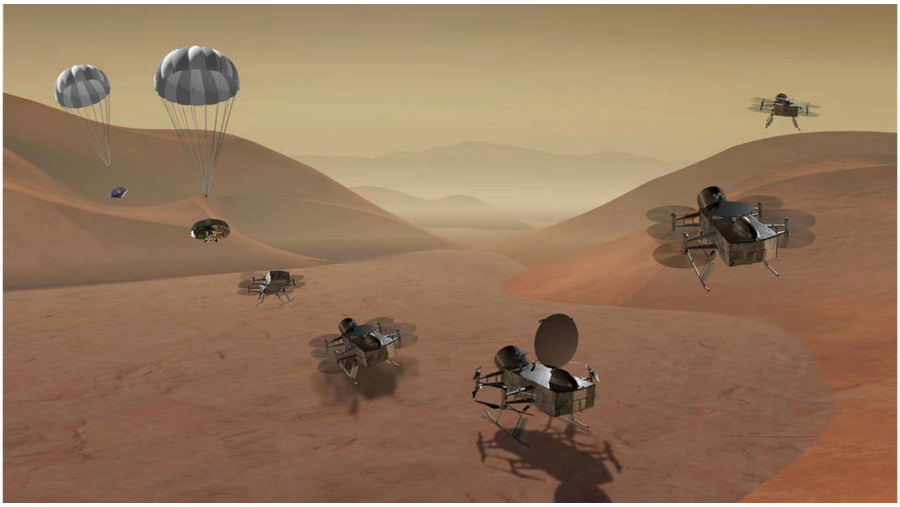

Figure 4.8. Dragonfly mission concept of entry, descent, landing, surface operations, and flight at Titan. Figure credit: NASA. https://www.nasa.gov/image-feature/dragonfly-dual-quadcopter-lander.

by nearly four decades ago. Even more exciting is the possibility of a Venus atmosphere sample-return mission of both gas atmosphere and cloud particles, to be analyzed on Earth with the most sophisticated instruments.

The Red Planet Mars is one of the main targets for the search for life in the Solar System, and several space agencies have plans to organize the first sample-return mission. This will be a unique opportunity to analyze samples from Mars—apart from Martian meteorites—with state-of-the-art equipment in the laboratory, in the hope that biosignatures will be unequivocally detected.

While it is not expected to also have sample-return missions to icy moons of our Solar System in the near future, analysis of their subsurface oceans has been performed indirectly by studying plumes that contained ice grains and vapor. In the next decades it would be ideal to develop technology in order to penetrate the surface ice and to directly analyze the subsurface ocean.

A next project is the launch of the future Dragonfly mission to Saturn's moon Titan (figure 4.8)—expected arrival 2036—that will assess its habitability conditions and prebiotic chemistry. Titan (figure 4.2) may become the main target of future life detection missions in 2050 and beyond. Potential microbial lifeforms living on Titan may use methane and other hydrocarbons present in the methane cycle [174, 175] rather than water, challenging our current ideas about the requirements for life in our Solar System.

4.2.2.2 Exoplanets: 2050 and beyond

Exoplanet astronomers are confident that the incredible pace of discovery during the last 10 years will continue into the next decade and beyond. New observatories with capabilities that have improved orders of magnitude with respect to their predecessors will open up new regions in parameter space, and computational power and tools such as AI also make huge progress.

In the search for signs of life on exoplanets, however, there are two inevitable limitations. First, despite the multitude of stars in the night sky and the hundreds of billions of stars in our own Galaxy, only stars in our solar neighbourhood (out to about 30–100 light years) are suitable targets for the search for planets with possible atmospheric biosignature gases. This is because exoplanets are adjacent to their much larger and much brighter host stars. This makes the signal of rocky exoplanet atmospheres so small that we require bright stars for observations. Because nearby stars are brighter than distant stars, our search is necessarily restricted to very nearby stars. In other words, the current generation of exoplanet astronomers is pinning its hope on the search for signs of life on what will be a very small number of habitable planets. If we do not find exoplanets with signs of life on the handful of exoplanets accessible to us in the next 10–20 years, the challenge will be how to scale up our efforts in 2050 and beyond, to access more, fainter, and more distant stars. Larger aperture space telescopes are limited to what fits within a rocket fairing, even for telescopes whose primary mirrors can be folded up and deployed on orbit. Larger apertures have been considered (e.g., the Large UV/Optical/Infrared Surveyor LUVOIR), and to design aperture sizes needed to reach even more distant stars, the next generation of astronomers will have to facilitate space-based manufacturing and space-based assembly.

The second inevitable limitation is a harsh reality: we may never know if an exoplanet atmospheric biosignature gas is indeed produced by life, or instead is produced by some unknown abiotic planetary chemistry. Indeed, this is a limitation difficult to overcome with remote sensing alone, leaving us with what we might call ‘inhabited candidates for exoplanetary life ’. As scientists, we are used to coping with uncertainty and unforeseen surprises, but when it comes to the search for life, would a *possible* detection of a sign of life be enough proof for scientists, and eventually for the taxpayer? If not, we need to consider brand new types of futuristic telescopes so that we might follow up in detail on any planets that show signs of life.

Ultimately, we will need to have a detailed look at any exoplanet that shows some evidence of hosting life. There are two paradigm-changing concepts under development. Both concepts, however, require a new operating timescale, at least 25 years from launch, before we get useful scientific results.

Breakthrough Starshot (https://breakthroughinitiatives.org/concept/3) is an effort funded by Breakthrough Initiatives to send thousands of tiny spacecrafts dubbed ‘starchips’ to Alpha Centauri B, a star in the nearest star system about four light years from Earth. Each starchip will deploy a solar sail of about 4 meter in diameter which will enable an acceleration to 20% of the speed of light, using the radiation pressure from a bank of coherent ground-based lasers with a combined power output of Gigawatts. The starchips that survive the 20-year journey will rapidly fly by Alpha Centauri B and any planets in the system. They will take images and relay them back to Earth, with a data transmission time at the speed of light equal to 4.4 years. The Starshot project has defined 19 technological challenges that are currently being worked out to make the mission possible during the next generation.

The Solar Gravitational Lens Telescope is an observatory that would operate very far from Earth at 550 AU (where 1 AU is the Earth–Sun separation), a distance to which no spacecraft has ever traveled yet. The concept is to use the Sun as a powerful gravitational lens with a focal distance of 550 AU [176–178]. The solar mass bends space such that any distant well-aligned background object will have its light rays bent towards its focal point. The background object then becomes magnified and spatially resolved. The better aligned the distant object is with the gravitational lens the higher the magnification and the better the achieved spatial resolution. Using the Sun as a gravitational lens this telescope could image exoplanetary surfaces at a resolution of 10 km. To reach 550 AU in a reasonable time, typically 25 years),the spacecraft would have to travel one order of magnitude faster than the Voyager spacecraft, the single one that has left our Solar System so far and that has taken about 45 years to reach a distance of 130 AU. Upon reaching the Sun's gravitational lens focal point, the spacecraft must follow the motion of the Sun-exoplanet focal line with a tailored trajectory in order to monitor the exoplanet. The planet image would be huge—on the order of kilometers—which is much larger than any conceivable telescope focal plane. For these and other reasons sophisticated computer algorithms would be needed to reconstruct the planet image. Because each planet has its own focal point with the Sun's gravitational lens, the telescope would be single-purpose, able to study only one exoplanet along the line of sight. Multiple telescopes could be sent into different directions, each for a pre-chosen target [178]. Despite the list of challenges, the revolutionary possibility for the direct imaging and spectroscopy of an Earth-like exoplanet up to 30 parsec away at a spatial resolution as small as 10 km makes this concept highly worth the investment.

4.2.2.3 Private sector

The number and kinds of opportunities to explore planets of the Solar System may increase with the growing involvement of private commercial space-flight companies. We have already mentioned the incredible utility of small space satellites in Earth orbit and a growing number of countries have become involved that were not traditionally space-faring nations. We get a glimpse into what could be a possible new future for Solar System exploration, that of small, focused missions visiting Solar System planets. Rocket Lab, a private company in the USA and New Zealand, has plans to launch their two-stage Electron rocket with the Photon satellite bus to Venus to drop a probe, most likely not before the end of 2024. The probe will be a light-weight (about 20 kg) with a small science instrument of around 1kg without parachute, spending just a few minutes in the Venus clouds but about an hour overall descending to the surface. Virgin Orbit's new LauncherOne, a rocket that is launched in the Earth's atmosphere from a modified Boeing 747 plane, can travel to Mars or Venus with an additional rocket stage to launch more mass than the current version. Public–private partnerships may become the norm in 2050 and beyond. Unlike national space agencies, private companies usually accept to take more risk at lower cost for novel space missions.

4.2.3 Conclusions and recommendations

The Grand Challenge of the search for life beyond Earth is ripe for huge progress. With strategic and concentrated international investment, we can accelerate the search, and increase our chance to find life beyond Earth, if it exists. Increased and highly focused efforts and funding in the following areas will serve this goal: field work in extreme environments on Earth that serve as planetary analogues; infrastructure and fundamental research for the study of habitability, life detection, and biosignatures; development of prototype equipment for future life detection missions; planetary protection development for both *in situ* measurements and for sample-return mission, shared data, experiment hardware, and protocols published in open access databases. It is also important to attract the next generation of scientists, with active public outreach and training of the next generation scientists and engineers.

The huge progress in technology and design that has taken place since the dawn of the space age only half a century ago means that we can send landers to the many targets in the Solar System that exhibit the necessary ingredients for life, and in some cases even return a sample to Earth. The discovery and characterization of exoplanets has come a long way in the new millennium since humans have pondered the mysteries of the multitude of stars. We are lucky to be the first generation that will not just imagine but that can truly carry out the search for extraterrestrial life. It is closer than it has ever been, and, once discovered, it will put an end once and for all of the often-believed unique position of Earth and change drastically our vision of the Universe.

Acknowledgments

Centro de Química Estrutural acknowledges the financial support of Fundação para a Ciência e Tecnologia (FCT) through projects UIDB/00100/2020 and UIDP/00100/2020. Institute of Molecular Sciences acknowledges the financial support of FCT through project LA/P/0056/2020. This work was partially supported by NASA Grant 80NSSC19K0471 and Heising-Simons Foundation Grant 2018-1104.

4.3 Artificial intelligence: powering the fourth industrial revolution

Angelo Cangelosi[1] and David Vernon[2]
[1]University of Manchester, Manchester, UK
[2]Carnegie Mellon University Africa, Kigali, Rwanda

Artificial intelligence (AI) is the branch of computer science and engineering that allows us to harness the power of computing and technology to mimic and extend human intelligence. Together with ubiquitous communications and near-universal access to information, AI is driving the Fourth Industrial Revolution, ushering in an era of unprecedented and rapid change in how humans live, work, and relate to one another through the fusion of physical, digital, and biological technologies. In this article, we trace the origin and evolution of the different strands of AI and consider the implications of its pervasive presence in society, addressing some of its many applications—in medicine, robotics, the World Wide Web and social media, and sport—and their impact on society across the globe, in developed and developing countries, and the ethical issues it raises for humankind.

4.3.1 What is AI, where did it come from, and where is it taking us?

In 1960, J C R Licklider predicted a symbiotic partnership between humans and computers that would perform intellectual operations much more effectively than humans alone can perform them [179]. Today that symbiotic partnership is being realized through AI, a technology that both amplifies and extends human cognitive abilities. AI forms the foundation of the fourth industrial revolution, a revolution that is characterized by a fusion of physical, digital, and biological technologies, powered by AI and enabled by ubiquitous communication and near-universal access to information. It is irreversibly altering how humans live, work, and relate to one another [180]. At the same time, it is important to consider how to harness AI within an ethical framework that achieves economic benefits and social development for all.

The world of the AI-powered fourth industrial revolution may well be the destination, but how did AI get started? For many people, the discipline of AI has its origins in a conference held at Dartmouth College, New Hampshire, in July and August 1956. It was attended by luminaries such as John McCarthy, who coined the term artificial intelligence, Marvin Minsky, Allen Newell, Herbert Simon, and Claude Shannon, all of whom had a very significant influence on the development of AI over the next half-century. The essential position of AI at this time was that intelligence—both biological and artificial—is achieved by computations performed on internal symbolic knowledge representations, an approach referred to as 'computationalism', grounded in cognitivist psychology, and often referred to as GOFAI: good old-fashioned artificial intelligence.

However, AI has other roots in cybernetics, which is concerned with self-organization, regulation, and control [181]. In 1950, Grey Walter developed two robotic turtles, Elmer and Elsie, that could roam around a room, find a charging station, and recharge themselves. These systems were built on behaviorist psychology by using associative and reinforcement learning in relatively simple neural networks, rather than focusing on internal models and symbolic computation. Neural networks process information by propagating it through an interconnected layered network of relatively simple processing units: artificial neurons, very simplified versions of the neurons in biological brains. This approach referred to as 'connectionism', progressed in parallel with the computationalist approach over the next 60 years and more. We'll say more about computationalist symbolic AI in section 4.3.2.1 and connectionist AI in section 4.3.2.2.

From the outset, symbolic AI was concerned with producing intelligent artifacts that exhibited the versatility, flexibility, and robustness of humans in rational problem-solving. For this reason, it became known as strong AI. Despite the early optimism, strong AI proved to be very difficult to achieve. Consequently, AI techniques began to be applied in more limited domains with stronger constraints and a narrower focus. This approach became known as weak AI. Despite continual progress in both symbolic AI and connectionist AI in the 1970s and 1980s, performance on more challenging problems was disappointing and the popularity of AI waned during a period known as the AI winter.

The AI winter came to an end in the 2000s when, building on research in the late 1990s, artificial neural networks with deeper network topologies (i.e., networks with many more layers than had been used in the mid-1980s to mid-1990s) and new learning techniques were introduced, leveraging the recent availability of much greater computing power in the form of graphic processor units (GPUs) and much larger datasets to train the networks. This period also saw the development of some landmark probabilistic approaches to AI, perhaps the most celebrated of which was the Watson system from IBM (named after its founder Thomas J Watson) which won the TV show *Jeopardy!* in 2011, beating two human champions in answering rich natural language questions over a very broad domain of topics. The success of Watson was the result of probabilistic knowledge engineering that integrated many knowledge sources and exploited many techniques for search, hypothesis formulation, and hypothesis evaluation [182].

4.3.2 The nature of AI

4.3.2.1 Symbolic AI and GOFAI

One of the key historical, methodological, and epistemological approaches to AI is that of 'Symbolic AI' (i.e., GOFAI). This has its origins in the 1950s (i.e., part of the 1956 Dartmouth Workshop for the start of the AI movement) and constituted the primary, classical approach in the first 30 years of AI research, before the first AI Winter and the advent of Connectionist AI and machine learning [183].

The term 'symbolic' refers to the fact that AI algorithms and programs are based on a set of symbols and symbol manipulation processes. Two of the founding fathers of symbolic AI, Allen Newell and Herbert Simon, proposed the concept of a **Physical Symbol System**, 'a machine that produces through time an evolving collection of symbol structures' [184, p 116]. These symbols are purely formal and meaningless entities, though in practice they are normally interpreted by the programmer with a particular semantic content such as words, numbers, pictures, actions, etc. The symbolic expressions are created using logic formalisms, such as propositional logic with Boolean connectives (e.g., 'Red AND Round') or predicate calculus (e.g., 'Apple(Red, Round)'). They can also be arranged in IF–THEN production rules (e.g., 'IF apple, THEN eat'). In specific symbolic systems such as semantic networks, each node has a symbol ('Red', 'Apple', 'Fruit') with links having a label for the semantic relationship between nodes (e.g., 'IS A' or 'HAS') and hierarchical relationship between nodes. A collection of symbolic structures for a specific domain constitutes the knowledge base used by the system to reason about the problem. In general, symbols systems solve problems by using the processes of heuristic search [184], where the search for the optimal link between the problem definition and its solution must be guided by heuristics (i.e., rules of thumb that help guide the program toward the solution in an optimal way). The AI heuristic search and planning algorithms are widely used today for scheduling and logistics, data mining, games, searching the web, and planning in robotics.

An important aspect of the GOFAI approach is the idea that symbolic systems can model human intelligence. Newell and Simon [184] proposed the **Physical Symbol Systems Hypothesis**, which states that 'A Physical Symbols Systems has the necessary and sufficient means for general intelligent action' [184, p 116]. This is why GOFAI systems have been applied to modeling mathematical reasoning, natural language processing, planning, game playing, etc. A classic example of a GOFAI system is an expert system (i.e., a program that represents the knowledge of the human expert in a specific domain, using a set of IF–THEN production rules, and which can be used to offer advice to non-experts or to provide solutions to experts). Beyond the historical examples of the first expert systems, such as Mycin to support medical doctors in the diagnosis and treatment of infectious diseases, today expert systems have been developed in a wide range of domains (commercial, education, medical, and military applications), with some capable of highly complex planning on the order of tens of thousands of search steps [185].

The major strengths of GOFAI are its abilities to model hierarchical and sequential tasks, such as language processing, problem-solving, and games, and to represent knowledge bases using propositional contents and inference processes.

Some limitations of GOFAI systems are that these AI programs are brittle and they cannot learn new knowledge. This, as well as the initial strong claims about the power of symbolic systems to deal with general intelligence and any problem domain, led to the first AI Winter in the 1980s, and to the subsequent developments of connectionist and machine learning approaches, which we cover in the next section. However, significant achievements of GOFAI include the widespread use of commercial expert systems and their essential role in the games industry (to control the intelligent behavior of the virtual agents) including the historical victory of the IBM Deep Blue system in 1997 in beating the chess world champion Gasparov, and IBM Watson's victory in 2011 over two human champions in the *Jeopardy!* TV game [185].

4.3.2.2 Connectionist AI: from perceptrons to deep neural networks

Connectionist AI differs from symbolic AI in that information is processed by propagating it through a large interconnected network of relatively simple processing elements, typically implemented as artificial neural networks. They use statistical properties rather than logical rules to analyze information. Although the term *connectionist model* is usually attributed to Feldman and Ballard [186], the roots of connectionism reach back well before the computational era, with connectionist principles evident in William James' 19th-century model of associative memory [187].

Neural networks also have strong foundations in physics, as many of the mathematics concepts on neuron modeling and computation come from physics principles. For example, the Ising model (also known as the Ising–Lenz model), a mathematical model of ferromagnetism in statistical mechanics, inspired a model of associative memory [188] that was popularized by John Hopfield's recurrent neural network: the Hopfield net [189]. Boltzmann machines are variants of Hopfield nets that use stochastic rather than deterministic weight update procedures to avoid problems with the network becoming trapped in non-optimal local minima during training [190]. In the future, the principles of quantum mechanics may provide the basis for efficient neural networks [191], in particular, and for quantum AI [192], in general.

The seminal paper by McCulloch and Pitts [193], 'A logical calculus immanent in nervous activity', is regarded as the foundation of artificial neural networks and connectionism. Connectionism advanced significantly in the late 1950s with the introduction of the perceptron [194] and with the introduction of the delta rule for supervised training [195]. However, perceptron networks suffered from a severe problem: no learning algorithm existed to allow the adjustment of the weights of the connections between input units and hidden units in networks with more than two layers (i.e., multi-layered perceptrons, MLPs). In 1969, Minsky and Papert [196] showed that these perceptrons can only be trained to solve linearly separable problems and can't be trained to solve more general problems. This had a very negative influence on neural network research for over a decade, marking the beginning of a decade-long winter for connectionist AI.

Perceptron-based neural networks underwent a strong resurgence in the mid-1980s with the introduction of the backpropagation algorithm [197], which had previously been derived independently by Paul Werbos [198], among others [199]. Backpropagation finally made it feasible to train MLPs, overcoming the restriction highlighted by Minsky and Papert [196], thereby enabling MLPs to learn solutions to complex problems that are not linearly separable. This was a major breakthrough in neural network and connectionist research.

By the early 2000s, the traditional neural network approach had fallen out of favor because effective training was limited to relatively small networks, both in terms of the number of layers and the number of units per layer, due to the lack of computational resources for training and the infeasible amount of time required to train large networks. However, in the late 1990s, significant breakthroughs in deep networks, such as long short-term memory (LSTM) by Hochreiter and Schmidhuber [200] and convolutional neural networks (CNNs) by LeCun *et al* [201], heralded a new era in connectionism, although it took another ten years before they were widely adopted because of the lack of sufficiently large datasets and sufficient computational power for training. A CNN network is similar in principle to the multi-layer perceptrons of the 1980s and early 1990s but they have more layers, each of which performs a different function. In a CNN, convolution refers to the application of a filter to the data being processed by the neural network. The key feature of a CNN is that these filters are learned by the network during the training phase. This marked a significant departure from previous approaches where the filters, and the features they extracted, were the result of hand-crafted design. Consequently, CNNs can map directly from the input space (e.g., the image to be classified or the image in which you want to search for a given object, directly to the image label or the object location). For this reason, they are referred to as end-to-end systems. The first CNN was created by Yann LeCun, focusing on handwritten character recognition [201]. In 2011, AlexNet [202], a CNN with seven hidden layers, won the ImageNet Large Scale Visual Recognition Challenge.

Since then, deep neural networks have been applied successfully in many challenging applications [203, 204]. The networks have become deeper, with 22 or many more layers, and performance has improved through the use of more effective activation functions (e.g., the rectified linear unit ReLU), the use of specialized layers (e.g., pooling), more advanced learning techniques (e.g., batch normalization and dropout), techniques to overcome the problem of vanishing gradients (where the error terms become too small to effect an improvement in network performance as they are propagated back in a deep network), and a better understanding of how to adjust the system hyper-parameters during training to improve performance.

While CNNs proved their mettle with a very impressive performance in image recognition, object detection and localization, face detection, face recognition, and object tracking, new forms of recurrent neural networks proved very successful on problems that involve processing and analyzing sequences of states (e.g., in natural language), by exploiting new recurrent elements such as long short-term memory (LSTM) and gated recurrent units (GRU).

Progress has continued, with modern architectures successfully combining the power of deep CNNs and LSTMs to address problems that involve both images and language (e.g., automatic image annotation and captioning, image retrieval, and synthesis based on linguistic descriptions) [205].

Progress using deep neural networks for language understanding and generation has recently been advanced even further with the series of generative pre-trained transformer (GPT) architectures, culminating, for now, in GPT-3 [206]. This system is capable of generating natural language text that is often indistinguishable from that generated by humans.

4.3.2.3 Statistical and machine learning

A parallel development in AI in the last 20 years, with partial overlap with the AI connectionist approach, has been that of machine learning. This is the field primarily based on a variety of **statistics-based inference methods** that use large datasets to estimate (i.e., learn) the parameters of a model that has classification and predictive capabilities. This approach developed in conjunction with AI research in computer vision and speech (or more generally, pattern recognition), in robotics (e.g., reinforcement learning), and in neural networks (MLP and deep neural networks). Some people today use the terms AI and machine learning interchangeably, especially because of the big, common emphasis on deep learning. But as we will see below, machine learning keeps a distinctive emphasis on data-driven statistical inference methods.

Amongst the various inferential strategies in statistics (e.g., analogical inference, domain-specific inference, and structural inference), the bulk of machine learning uses the structural inference approach. This uses domain-general algorithms which exploit the **internal structure of the data**, rather than identifying the semantic, domain-specific, content of the data. Structural inference is the basis of most machine learning frameworks, such as the well-known methods of regression, neural networks, and Bayesian networks [207]. Given this data-centric (sometimes known as 'data-hungry') approach, the recent, easy availability of potentially unlimited data from social media and the web, and wider access to cloud-based parallel computing systems such as GPUs (which are necessary to apply computationally intensive statistical computations on large datasets) can in great part explain the recent, impressive contribution of machine learning to AI, and information technology in general.

Machine learning comprises a set of methods typically grouped into supervised and unsupervised techniques, as well as reinforcement methods. **Supervised learning** algorithms need a labeled dataset (i.e., where each data point such as an image of a dog, for example) is associated with a supervision signal or ground-truth (e.g., the category label 'dog'). The learning algorithm has to find the parameters of the model (e.g., weights of a neural network) using the error between the model's guess and the supervision label. Examples of supervised learning algorithms include MLP, CNN, and LSTM neural networks, decision trees, support vector machines, and regression. **Reinforcement learning** can be considered part of the supervised approach, but where the supervision to learn a policy (e.g., actions that should be taken when certain sensory conditions prevail) is guided by a reward function. **Unsupervised learning** algorithms do not require a labeled dataset, as they discover the regularity in the data and their organisation in separate categories. Examples of unsupervised learning include clustering algorithms such as k-means and autoencoder neural networks,

An important set of machine learning approaches is that of Bayesian learning algorithms. The general Bayesian framework is based on the intuition that the beliefs after observing some data are determined by the probability (prior probability distribution) of each possible explanation given that data. When processing a dataset, the machine learning algorithm uses the Bayesian rule to calculate the correct probability distribution over the hypotheses given that data. And given large datasets, the computations required for Bayesian learning become too difficult to be done analytically, thus the recent boost of Bayesian algorithms with easy access to parallel computational resources [207].

Machine learning in large part is responsible for the recent successful developments of AI. The most successful and widely used applications in speech recognition, computer vision, natural language processing, and robotics applications are based on deep learning and other Bayesian approaches. This includes the design of DeepMind's AlphaGo and AlphaZero systems, based on the combination of deep neural networks, reinforcement learning, and AI search algorithms, which, as we will see in section 4.3.4, were able to outperform human champions in the Go game, even without human knowledge or supervision [208].

4.3.3 Example applications

4.3.3.1 AI applications in medicine

The application of AI to medicine and healthcare has its origins in the early GOFAI developments of expert systems, such as the MYCIN for infectious diseases and DENDRAL on the discovery of chemical compounds. More recently, the advent of machine learning and its focus on learning from data has led to a resurgence of the development of medicine AI systems.

Deep learning methods, such as CNNs because of their impressive performance with 2D image recognition, have been widely used for **image-based cancer detection and diagnosis** [209]. For example, in skin cancer diagnosis, the performance of CNNs to classify biopsy-proven clinical images (e.g., malignant melanomas versus benign nevi) was on par with that of 21 board-certified dermatologists. A recent landmark achievement in medicine and biochemistry is the **AlphaFold AI model for protein folding**. This was developed by Google DeepMind and was the winner in 2020 of the biennial critical assessment of protein structure prediction (CASP) competition. AlphaFold achieved a performance similar to the results from experimental methods.

Medical AI applications bear significant **technological and ethical challenges**. One key issue is the reliance on the quality and variety of the training data, as healthcare datasets typically are sparse, noisy, heterogeneous, and time-dependent. Moreover, new methods and tools are needed to enable interactive machine learning to interface with healthcare information workflows, keeping the human in the loop [210, 211]. There are also important ethical considerations. For example, the need for explainable systems so that clinicians (both novice and expert doctors) can access causal explanations of the AI's decision-making process [212]. We return to this issue of trust and explainability in AI in section 4.3.1.

4.3.3.2 AI applications in robotics

AI is used in robotics for many purposes, including autonomous navigation, task planning, task execution, object detection, object grasping and manipulation, inspection and surveillance, social human–robot interaction, including natural language processing, facial recognition, sentiment analysis, gesture understanding, and intention recognition. It is also used in an extensive range of robots. At the time of writing, the IEEE robots website [213] features 229 robots of many different types: wheeled, legged, tracked, airborne, underwater, and humanoid, targeting consumers, entertainment, education, research, medicine and healthcare, disaster response, service and industrial, aerospace, military and security applications, telepresence, self-driving cars, and agriculture. Perhaps the epitome of AI in robotics is the goal of creating a collaborative robot (i.e., one that can share a common goal and share the human's intentions to achieve that goal, acting jointly with the human, paying attention to what the human is doing, and, crucially, anticipating any help the human might need to complete whatever tasks she or he is working on).

One example of AI in robotics is robot-enhanced therapy [214] where **robots assist a psychotherapist working with children** with autism spectrum disorder. Under the guidance of clinical practitioners, this project developed interactive capabilities for social robots that allowed them to engage a child in clinically derived exercises. The robot can operate autonomously for limited periods under the supervision of a psychotherapist. AI plays a major part in the success of this application, specifically in its cognitive ability to interpret body movement and appearance-based cues of emotion. This allows the robot to assess the child's actions by learning to map them to therapist-specific classes of behavior. In turn, the robot also learns to map these child behaviors to appropriate robot responses, again as specified by the therapists.

4.3.3.3 AI applications in the web and social media

AI is having a tremendous impact on a variety of applications and functionalities for the web (e.g., search algorithms, music and video recommendations, automatic translation) and social media (e.g., news selection and recommendation, sentiment analysis, face recognition). Although this progress is resulting in clear benefits to people and society, it also carries important ethical considerations and risks.

AI has significantly changed the **search algorithms** for the web. For example, Google's initial PageRank algorithm (based on standard mathematical methods) has now developed a collection of search tools, such as the Hummingbird framework. This framework complements PageRank's results with RankBrain, based on machine learning algorithms for entity recognition, and the recently introduced BERT (Bidirectional Encoder Representations from Transformer), which uses a neural network for natural language processing.

Another example of the widespread use of AI and machine learning algorithms on the web is for **recommender systems, which make** recommendations for related purchases on e-commerce sites, suggestions of related news and friends in social media, and personalised recommendations in media-streaming sites and apps. In fact, 80% of movies watched on Netflix are based on AI recommendations [215]. As in other domains, deep machine learning systems have become the default algorithm.

AI applications for **face recognition** have also become widespread on the web and in social media. These algorithms can be used for image matching and people recognition (e.g., in social media photo tagging) as well as for authentication (e.g., to implement secure access in some smartphone systems). However, face recognition algorithms based on learning from datasets have important ethical implications (e.g., regarding possible biases in the data used for the training). For example, in 2018, a seminal paper by computer scientists Buolamwini and Gebru [216] demonstrated that leading facial recognition systems produced substantial disparities in the accuracy of gender classification (e.g., with error rates of up to 34.7% in the classification of darker-skinned females, whilst the maximum error rate for lighter-skinned males was 0.8%). This highlights the urgent need to address and remedy implicit bias in such systems and make sure they are based on fair, transparent, and accountable facial analysis algorithms.

4.3.3.4 AI applications in sports

While the use of statistical analysis is well-established in sports, AI is taking it to a new level. Possible applications range across the entire spectrum of activities. Barlow and Sriskandarajah [217] identified eleven applications across 17 sports that are being or will be impacted by AI. These applications include identifying talent and determining optical game strategies.

AI technologies such as computer vision have been used routinely to **assist with umpiring during games**, especially using automated ball tracking and line calling applications. For example, the Hawk-Eye system [218] visually tracks the trajectory of the ball using six high-speed cameras, the images from which are used to triangulate the ball's position over time, displaying a virtual reality trajectory of its statistically most likely path. AI can also be used for **automated generation of video highlights**, while integrated vision and natural language technology can be used for automated generation of copy for publication in print and online.

The All England Lawn Tennis Club hosts the annual tennis championship at Wimbledon and uses IBM's Watson technology to provide a variety of services, including **real-time match reports and uncovering player insights**. It also powers a voice-activated cognitive assistant 'Fred', named after the late champion Fred Perry, to help spectators find their way around the venue [219].

To identify successful game strategies, an **AI system can play against itself**, as the DeepMind AlphaGo system did, before going on to beat Lee Sedol, the winner of 18 world titles, in 2016, and achieve 60 straight wins in time-control games against top international players in 2017 [220]. Subsequently, in AlphaGo Zero, even better performance was achieved based purely on reinforcement learning without any prior supervised training. Apart from its formidable performance, what is significant about AlphaGo is that it uncovered several innovative strategies that greatly surprised expert players, demonstrating the potential for AI to augment human abilities and exceed human performance.

4.3.4 Future challenges

4.3.4.1 Collaborating with machines and robots

AI has contributed significantly to the design of intelligent control models and cognitive architectures for sensorimotor behavior (e.g., perception, navigation manipulation) and cognitive capabilities (e.g., planning, language) in robots, as we have seen in section 4.3.1. But major challenges still remain, specifically with regards to the robot being able to handle the complexity of real-world scenarios (i.e., cluttered, dynamic, unpredictable environments where objects to be grasped or obstacles to be avoided are difficult to see, can be occluded, or change their position over time). Beyond the complexity of designing skills in individual robots, a significant challenge for robots, and intelligent machines in general, is that of handling **interaction with people for collaborative tasks**, also known as human–robot interaction (HRI) or social robotics [221]. This type of interaction includes a variety of scenarios, such as joint action in a flexible manufacturing setup between a worker and a cobot (i.e., a collaborative robot), assistive robot companions for older and disabled people or in hospitals and care homes, and robot tutors for education or entertainment.

The research challenges on the use of AI for the design of social and cognitive skills for interaction include, for example, the **capability of 'intention reading' and the implementation of an artificial 'Theory of Mind'** [222]. Intention reading is the capability of the robot to detect the human user's intended goal of the joint interaction. The 'Theory of Mind' describes a more general view of intention reading, as this concerns the robot's capability to understand and predict the beliefs, desires, and goals of the person; as shown in figure 4.9. AI methods, such as Bayesian networks and deep learning, can be used to build artificial Theory-of-Mind skills in robots [222].

Figure 4.9. This sequence of pictures depicts a situation in which the iCub humanoid robot (www.icub.org) is interacting with a human, reading her intention to get her phone from her bag, and alerting her to the fact that it is on the desk, hidden from her by the laptop. This sequence has been staged to illustrate the future capabilities of a cognitive robot and has not yet been implemented. Reprinted from [223], copyright (2021) with permission by Springer Nature.

Another important future research direction in AI for collaborative robotics concerns the quality of the interaction (i.e., the design of **long-term and trustworthy interaction and well-being in human–robot collaboration)**. Long-term interaction requires the robot to be able to engage in continuous, meaningful, and contextualised interaction over a series of interactions lasting for days, weeks, or longer. This will require the ability to recognize the person and their personality and preferences, to remember recent interactions, and to engage in empathic behavior with the person's needs [224]. Trustworthy interaction, a growing field of research, requires people's acceptance and trust of the robot's behavior and decision-making process. This is also linked to ethical issues regarding explainable AI (see section 4.3.4.3) and to the achievement of people's and robot's reciprocal theory of mind [225, 226].

4.3.4.2 Self-learning and self-programming machines

The quest for the automatic generation of programs and codes, also known as program synthesis or self-programming machines, has been one of the main challenges of AI since the outset. With the advent of machine learning approaches, AI has started to put emphasis on self-learning machines, which can learn with no or minimal supervision from humans.

This **self-programming machine** challenge has only recently received a significant boost through the combination of deep learning and NLP methods. For example, DeepCode is a code generator that uses a neural network to predict the properties of the program that can produce the outputs given specific inputs [227]. SketchAdapt is a software system that learns, without direct supervision, when to rely dynamically on pattern recognition and when to perform symbolic search for explicit reasoning [228].

Recently, the GPT-3 system [206] has been proposed for natural language generation, with the potential application to automatic program synthesis. GPT-3 is a large-scale deep learning model for natural language processing with an order of magnitude more parameters than any previous NLP model. The model can generate new text without the need for further training or task-specific fine-tuning of its parameters. GPT-3 can produce samples of news articles that human evaluators have difficulty distinguishing from articles written by humans. In addition, GPT-3 has been used for generating programs, such as the code to create the Google homepage [229].

Regarding the challenge of creating a **self-learning machine**, the first attempts to design AI systems and robots that autonomously learn without supervision from humans have recently been realized in developmental robotics (also known as autonomous mental development). This area of robotics takes inspiration from child development to design robots that go through stages of development for the incremental acquisition of sensorimotor and cognitive skills [230]. Another example of self-learning AI is the AlphaZero system mentioned earlier in which artificial agents play the game Go against each other, bootstrapping their final learning capabilities. This led to the acquisition of skills that far outperformed the skills of the best human players ([231]; see also section 4.3.3.4).

4.3.4.3 *Social and ethical aspects of AI*

AI can bring significant benefits to all. However, the examples we have given so far focused on applications in the developed world and, indeed most of the national strategies on AI have been created by governments in developed countries [232]. Nevertheless, the fourth industrial revolution in general, and AI in particular, is just as relevant for developing countries. For example, AI is having an increasingly positive impact in Africa, in sectors such as energy, healthcare, agriculture, public services, and financial services [233, 234]. It has the potential to drive economic growth, development, and democratization, to reduce poverty, improve education, support healthcare delivery, increase food production, improve the capacity of existing road infrastructure by increasing traffic flow, improve public services, and improve the quality of life of people with disabilities [235]. The challenge is to ensure that developing countries have access to the technology and the datasets necessary for innovation in a socially and culturally acceptable manner (i.e., the democratization of AI, and open access to AI technology by developers everywhere). It is crucially important that the fourth industrial revolution, powered by AI, occurs in a fair, ethical manner [236].

AI can also be used for negative purposes, either intentionally or unintentionally (e.g., by fomenting religious, ethnic, social, and political divisions through fake misinformation created by deep networks) [237]. Of particular concern is the issue of implicit and explicit bias in the data that are used to train the AI models, thereby resulting in discrimination against people based on gender or race. Examples of bias against dark-skinned people include facial analysis [216], pedestrian detection [238], and predicting recidivism [239]. The success of AI in the future will depend on the elimination of such bias.

4.3.4.4 *Intelligence, brains, and consciousness*

Why does intelligence matter? Indeed, what is intelligence? Can a robot be conscious? Let's start by answering the question: what is intelligence? There are many possible answers but the one that has the most appeal derives from the answer to a different question: why do we have brains? The neuroscientist Daniel Wolpert provides an unexpected but compelling answer. He argues that we have brains to allow us to control movement [240]. This mirrors what Francisco Varela and Umberto Maturana say about cognition: 'Cognition is effective action' [241]. From this perspective, we see intelligence as the way to be effective in our control of our movements and in the way we act in the world. The key to understanding why this is so important—and so difficult—is to see that the number of possible ways we can move and act, and the number of possible outcomes of these movements and actions, is infeasibly large if we are to consider all the possibilities and choose the best one, or even a good one. This is what Allen Newell and Herbert Simon pointed out in their Turing Award (the equivalent of the Nobel Prize for computer science) lecture: 'The task of intelligence, then, is to avert the ever-present threat of the exponential explosion of search' [184] (i.e., the search for good ways to act). Newell and Simon were referring to the search for the solution to a problem, but it amounts to the same thing. This is a satisfyingly straightforward and very practical way of understanding intelligence and the brains that give rise to intelligence.

However, brains are even better than that. They also predict the need to act and the outcome of those actions, and they do so all the time, at every instant, as we act and as we anticipate the future, milliseconds ahead, seconds ahead, hours, days, years. Indeed, it has been argued that brains are, in effect, probabilistic prediction machines, meaning that they can deal effectively with uncertainty [242].

But what then of consciousness? Can robots and intelligent machines also be conscious? Can we take AI even further and build machines with artificial consciousness? Many people think that this is a distinct possibility. Indeed, according to Paul Verschure, 'understanding the nature of consciousness is one of the grand outstanding scientific challenges' and he proposes a scientifically grounded approach to addressing the challenge of answering the question of what consciousness is and how physical systems can develop it [243].

4.3.5 Summary and conclusion

AI impacts all aspects of human activity: it automates tasks, assists with decision-making, augments and extends our cognitive capabilities, and it can even operate autonomously, if we allow it, without recourse to human oversight.

AI began as an attempt to understand and replicate human intelligence, initially taking two routes to that goal, one via connectionism and one via symbolic computationalism, reflecting their inspiration from behaviorist and constructivist psychology, respectively. These two approaches waxed and waned in their own respective ways over the decades, to be joined in the 1980s by machine learning and in the 1990s by statistical machine learning, probabilistic inference networks, and other established disciplines in computer science. Breakthroughs in deep neural network learning and deep neural network topologies, aided by very large datasets and equally large increases in processing power, yielded great success in many application domains. The symbolic knowledge representation and reasoning approach also developed rapidly, especially in cognitive architectures, as knowledge bases and ontologies increased greatly in size and sophistication and as the hybrid paradigm, combining symbolic approaches and sub-symbolic connectionist approaches, was developed (e.g., in cognitive architectures such as Soar [244], ACT-R [245], and CLARION [246], among others).

While the success of statistical machine learning in narrowly targeted applications yielded great success, it did so at the expense of losing focus on AI's original goal of understanding and replicating human-level intelligence. There has been a resurgence of interest in what is now known as Artificial General Intelligence in cognitive science and cognitive systems. Still, the ultimate goal of replicating the versatility of human cognition remains elusive and it is unclear when it will be achieved. What is certain is that the AI quest will continue and AI in its many guises will continue to permeate our lives and change them, hopefully, for the better.

In seeking to steer the path to the future, it is likely that other strands of thinking will be woven into the fabric of AI, especially concerning the trustworthiness of AI in autonomous systems (i.e., its role in serving the bigger agenda of creating self-maintaining systems that can operate robustly and prospectively in the face of uncertainty and that can continually develop through self-programming as they interact with and learn from the world and the people in it). While there is much important work yet to be done to promote the development of democratized, trustworthy, ethical AI in the developed and developing worlds, an equal challenge will be how to control the role of AI in autonomous systems, possibly conscious ones, where the relationship with humans is no longer symbiotic. We are far from that point at present. but it is likely we will reach it, and everything will change quickly when we do.

In Ernest Hemingway's novel *The Sun Also Rises* there is a dialog between two characters which goes as follows. 'How did you go bankrupt?' Bill asked. 'Two ways,' Mike said. 'Gradually and then suddenly'. And so it will be with autonomous AI. Our collective responsibility is to work together in a directed manner during the present gradual phase so that, when the full impact of AI is suddenly felt, it will be for the greater good of all humankind.

Acknowledgments

In the preparation of this section, Angelo Cangelosi's work was partially supported by the Air Force Office of Scientific Research, USAF under Award No. FA9550-19-1-7002, by the UKRI TAS Node on Trust (EP/V026682/1), and the H2020 projects PERSEO, TRAINCREASE, and eLADDA.

4.4 Artificial life: sustainable self-replicating systems

Carlos Gershenson[1] and Jitka Čejková[2]
[1]Universidad Nacional Autónoma de México, Mexico City, Mexico
[2]University of Chemistry and Technology, Prague, Czechia

4.4.1 General overview

Nature has found one method of organizing living matter, but maybe other options exist—not yet discovered—on how to create life. To study life 'as it could be' is the objective of an interdisciplinary field called Artificial Life (commonly abbreviated as ALife) [247–249]. The word 'artificial' refers to the fact that humans are involved in the creation process. The artificial lifeforms might be completely unlike natural forms of life, with different chemical compositions, and even computer programs exhibiting life-like behaviours.

ALife was established at the first 'Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems' in Los Alamos in 1987 by Christopher G Langton [250]. ALife is a radically interdisciplinary field that contains biologists, computer scientists, physicists, physicians, chemists, engineers, roboticists, philosophers, artists, and representatives from many other disciplines. There are several approaches to defining ALife research. One can discriminate between soft, hard, and wet ALife (figure 4.10). '**Soft**' ALife aims to create simulations or other purely digital constructions exhibiting life-like behaviour. '**Hard**' ALife is related to robotics and implements life-like systems in hardware made mainly from silicon, steel, and plastic. '**Wet**' ALife uses all kinds of chemicals to synthesize life-like systems in the laboratory.

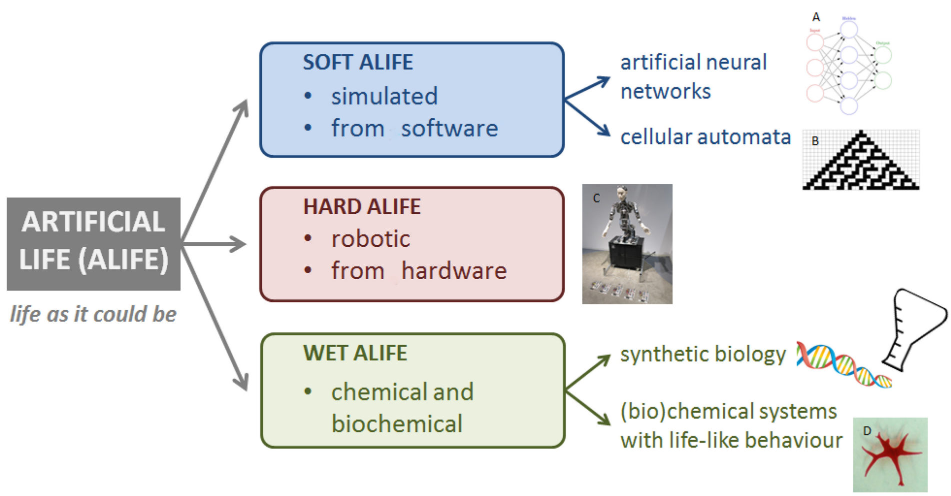

Figure 4.10. Artificial life research. (A) Artificial neural networks. (B) Cellular automaton. (C) Robot Alter 2. (D) Shape-changing decanol droplet.

Bedau *et al* [251] proposed 14 open problems in ALife in the year 2000, but none of them have been solved yet. Aguilar *et al* [249] summarized the ALife research challenges and divided them into 13 themes: origins of life, autonomy, self-organization, adaptation (evolution, development, and learning), ecology, artificial societies, behaviour, computational biology, artificial chemistries, information, living technology, art, and philosophy.

4.4.2 Soft ALife

Mathematical and computational models of living systems are naturally more abstract than robotic or chemical systems. Also, they are easier to build, so actually most ALife models are 'soft'. Still, we can discriminate between 'abstract' and 'grounded' soft ALife models. This distinction is more gradual than categorical. More abstract models focus less on physics or biology, and more on information and organization, while more physical models consider to a greater degree the actual components of living systems.

Some of the most abstract models include cellular automata [252–258], random Boolean networks [259–262] (figure 4.11), and boids [263], where there are basically space, time, and simple dynamics that may lead to complex behavior. Other abstract models have focussed on studying self-replication [264] or evolution [265–269]. Abstractly exploring the theoretical space of possibilities for living systems (necessary and sufficient conditions) has also been made at the 'chemical' level, with artificial chemistries [270, 271] and swarm chemistry [272] (figure 4.11). Another more 'grounded' strand of soft ALife involves the simulation of environments with realistic physics, either to evolve 'creatures' [273] or controllers for physical robots [274–276].

4.4.3 Hard ALife

Robots are physically situated [277, 278] and embedded [279] in their environment, so they have been useful to explore aspects of life related to behavior [280], traditionally studied by ethology [281]. There are also several soft ALife models of adaptive behavior [282, 283], dealing with physics (time, motion, inertia, gravity, hardware imperfections, etc) that already are an important challenge for ALife. As mentioned earlier, robotic controllers have usually been evolved in software that has been uploaded into hardware. This is also known as 'evolutionary robotics' [284–286]. Hard ALife models have also been used extensively to study the emergence of collective behaviors [287–291] (figure 4.12).

4.4.4 Wet ALife

Wet ALife is related to the effort of creating artificial cells in the laboratory from chemical and biochemical precursors. Living cells are the basic structural, functional, and biological units of all known living organisms. They are found in nature and produced and maintained by homeostasis, self-reproduction, and evolution [292]. In contrast, artificial cells (synthetic cells) are prepared by humans and only mimic some of the properties, functionalities, or processes of natural cells.

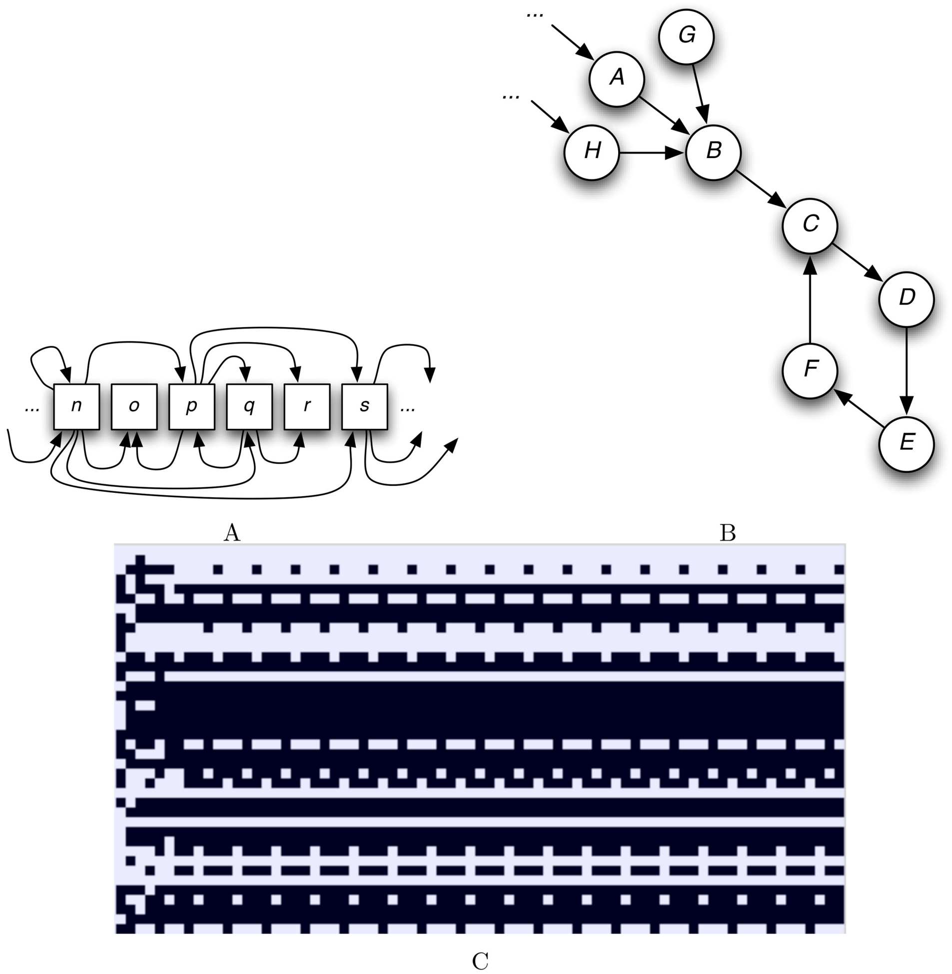

Figure 4.11. An example of a random Boolean network (RBN) [262]. (A) A structural network is formed by N Boolean nodes (can take values of zero or one) that are connected to K inputs randomly. The future state $(t + 1)$ of each node is determined by the current state (t) of its inputs following lookup tables that are also generated randomly (and then remain fixed). (B) The structural network defines a state transition network with 2^N nodes. Each state has precisely one successor, but it can have several or no predecessors. Thus, it is a dissipative system. Eventually, a visited state is repeated, and thus the network has reached an attractor. (C) Example dynamics of an RBN with $N = 40$ and $K = 2$, time flowing to the right. A random initial state converges into an attractor of period 4. A single RBN can have several attractors of different periods.

Although many laboratories are working on this task, the successful preparation of artificial cells having all features of natural living cells is still challenging. An artificial cell that can self-produce and maintain itself (a so-called autopoietic system) has not yet been demonstrated. All published papers with ‘artificial cell’ in the title describe usually simple particles that have at least one property in common with living cells. Nevertheless, a big challenge exists to synthesize an artificial cell having at least several properties shared by living cells. For example,

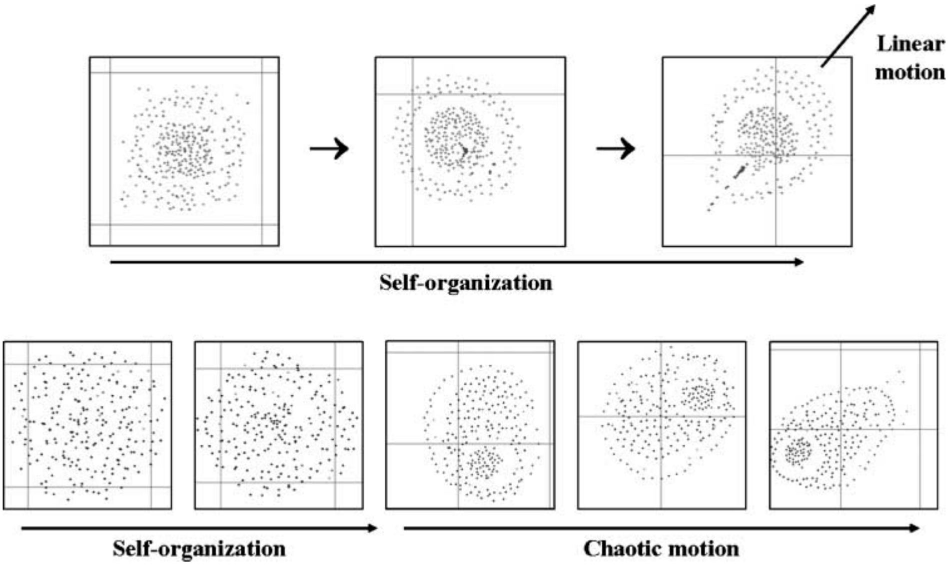

Figure 4.12. Examples of complex structures made of several different chemical species, designed using Sayama's interactive simulator [272]. The swarms self-organize from initially random states to a shape that looks like a horseshoe crab (top), or a biological cell-like structure that shows active chaotic movement after self-organization (bottom).

(i) the presence of a stable semi-permeable *membrane* that mediates the exchange of molecules, energy, and information between internal content and external environment while preserving specific identity; (ii) the possibility to sustain itself by using energy from its environment to manufacture at least some components from resources in the environment using *metabolism;* and (iii) the capacity of growth and self-replication including genetic *information*.

A simpler and less problematic approach is to synthesize protocells that are not necessarily alive, but that exhibit only a few life-like properties [293–295]. Protocells can be defined as simplified systems that mimic one or many of the morphological and functional characteristics of biological cells. Their structure and organization are usually very simple and can be orthogonal to any known living system. Protocells are used both to model artificial life and to model the origins of life. In the latter case, protocells are often considered hypothetical precursors of the first natural living cells.

In principle, two main motivations exist to create an artificial cell. One group of researchers aims to answer questions about the origin of life: they synthesize primitive cells which consist of a protocell membrane that defines a spatially localized compartment, and of genetics polymers that allow for the replication and inheritance of functional information. The aim is to create self-replicating micelles or vesicles and to observe the spontaneous Darwinian evolution of protocells in the laboratory [296, 297]. Other researchers want to prepare particles with life-like properties that mimic the behaviour of living cells, though without the ability to self-replicate or evolve. Such objects can move in their environment [298, 299], selectively exchange molecules with their surroundings in response to a local change in temperature or concentration, chemically process those molecules and either accumulate or release the product, change their shape [300, 301], and behave collectively [302, 303]. Such synthetic objects can be used for instance as smart drug delivery vehicles that release medicine *in situ*. These artificial cells could also be called chemical or liquid robots [304] (figure 4.13).

Recently, so-called Xenobots were introduced [305, 306]. Although they do not belong to traditional wet artificial life research, they should be mentioned here as a new approach on how to synthesise artificial organisms in the laboratory by using ALife tools. Xenobots were designed by an evolutionary computer algorithm and then assembled from embryonic frog cells *Xenopus laevis* (hence the name Xenobots). Whether Xenobots are 'living robots', 'living machines', or 'man-made animals' it remains a debatable question. Nevertheless, these creatures, smaller than

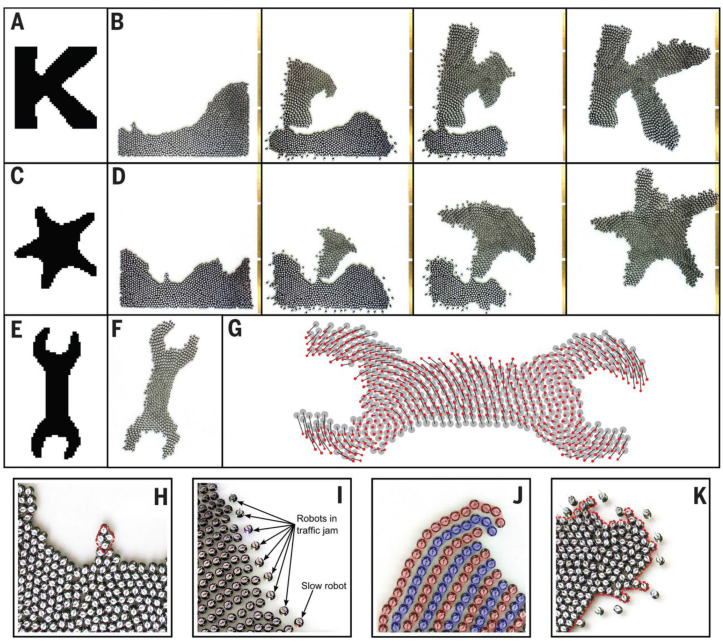

Figure 4.13. Self-assembly experiments using up to 1024 physical robots. Reprinted from [290] with permission from AAAS. (A, C, and E) Desired shapes are provided to robots as part of their program. (B and D) Self-assembly from initial starting positions of robots (left) to final self-assembled shape (right). Robots are capable of forming any simply connected shape, subject to a few constraints to allow edge-following (19). (F) Completed assembly showing global warping of the shape due to individual robot errors. (G) Accuracy of shape formation is measured by comparing the true positions of each robot (red) and each robot's internal localized position (gray). (H–K) Close-up images of starting seed robots (H), traffic backup due to a slowly moving robot (I), banded patterns of robots with equal gradient values after joining the shape (robots in each highlighted row have the same gradient value) (J), and a complex boundary formed in the initial group (dashed red line) due to erosion caused by imprecise edge-following (K).

a millimeter in size, can find potential applications in areas such as environmental remediation. At this moment, Xenobots have the ability neither to self-replicate nor to evolve, but the lifespan and the hypothetical ability to reproduce could be assessed and regulated in the future in accordance with ethical principles (figure 4.14).

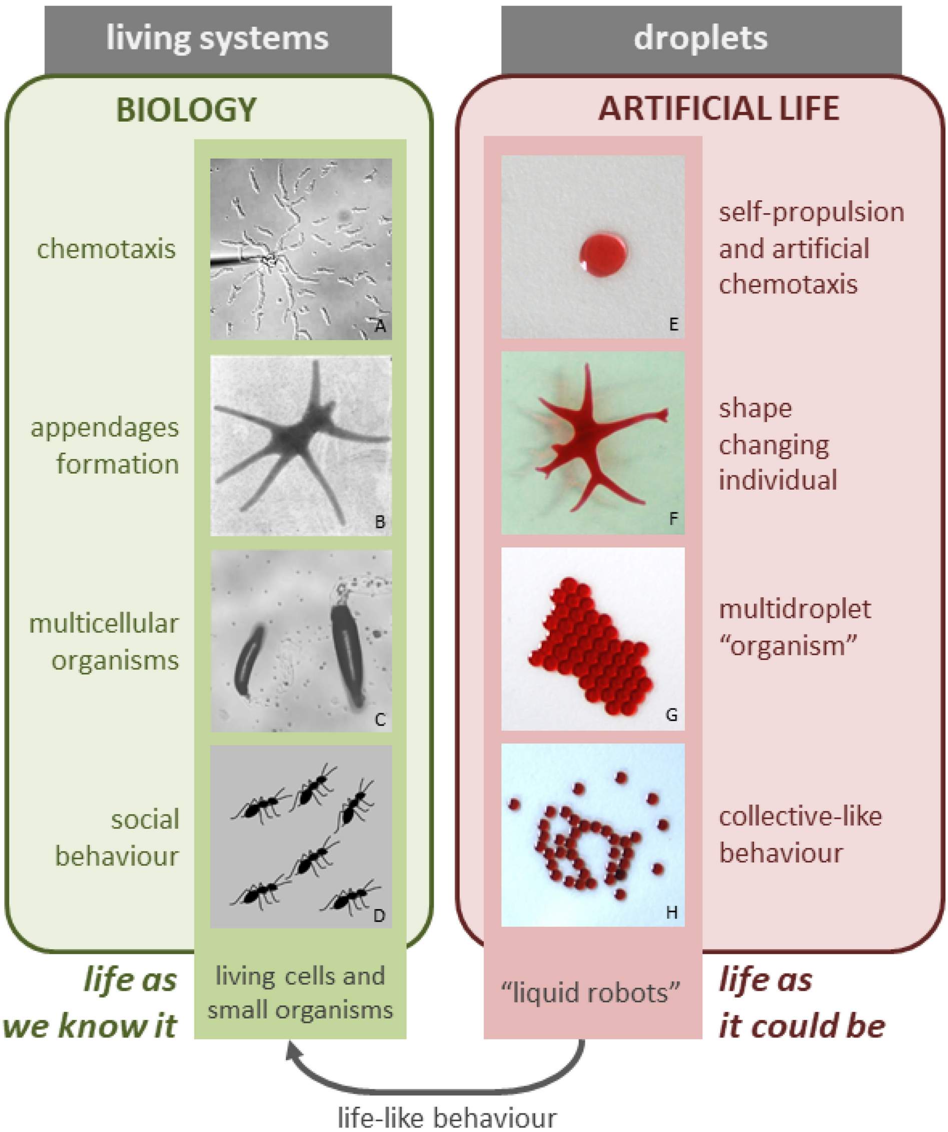

Figure 4.14. Schematic comparison of wet artificial life research on droplets to biology and studies of living systems. (A) Chemotaxis of *Dictyostelium* cells. (B) Prosthecate freshwater bacteria. (C) Multicellular slug stage of the *Dictyostelium* developmental cycle. (D) Schematic representation of ants group. (E–H) Decanol droplets were placed into an aqueous solution of sodium decanoate.

4.4.5 Self-replicating systems

Let us imagine a factory where robots are constructing other robots. We know such scenes from the movie *I, Robot* where the company U.S. Robotics produces humanoid robots, or from an island factory in the theatre play *R.U.R.—Rossum's Universal Robots*. However, these examples are fictional, and such factories have never existed in the real world and still belong to science fiction. Still, the idea of machines that beget machines was already in the air four centuries ago. A comprehensive study of the history and state of the art of self-replicating machines from a scientific and technological point of view can be found in a recent book by Taylor and Dorin [307].

Self-replicating systems can be categorized as follows: 'Standard-replicators' that could reproduce by building a copy of themselves; machines that are able not only to reproduce, but also evolve by natural selection as their living counterparts, are categorized as 'evolvable self-replicators' (evo-replicators); and so-called 'manufacturing self-replicators' (maker-replicators) have the ability not only to self-replicate, but also to create specific goods and materials as by-products when they self-replicate.

4.4.6 Challenges and opportunities: a 2050 vision

ALife will have to tackle very specific challenges in science and engineering: (a) Get the agents out of the lab and into the wild. (b) Make them able to interact with living organisms in a predictable way. (c) Make them sustainable and degradable in order not to deteriorate the already transformed environments. (d) Allow long-term operations of these agents, or even multi-generational timespans by self-reproduction of the agents. The development of sustainable technologies is thus urgently needed, and ALife could in part address that. The ALife systems should be characterized by robustness, autonomy, energy efficiency, materials recycling, local intelligence, self-repair, adaptation, self-replication, and evolution, all properties that traditional technologies lack, but living systems possess [308, 309].

Concerning **soft ALife**, perhaps one of the greatest challenges is that of open-ended evolution (OEE) [310–313]: can a program produce ever-increasing complexity (as it seems natural evolution does)? Hernández-Orozco *et al* [314] proposed that undecidability and irreducibility are requirements for OEE. It might be that the difficulty of achieving OEE is related to the limits of formal systems [315–317]. Simplifying the situation: formal systems cannot change their own axioms. This is a necessary condition for traditional logic, mathematics, and computation, but perhaps OEE requires precisely the possibility of changing axioms.

As for **hard ALife**, robots are becoming more and more sophisticated [318]. They are also gaining autonomy. However, as with the rest of artificial intelligence, all robots are specialists. They are good at performing the tasks they were designed for, but they cannot generalize and perform other activities. For example, a vacuum robot cannot paint. Not only because it lacks the appropriate hardware, but also because the software is task-specific. The so-called 'artificial general intelligence' has so far produced not more than mere speculations. Could it be that the limits of formal systems just mentioned for OEE also affect artificial intelligence in general and robots in particular? If so, can we find an alternative, to build robots that are not based on formal systems?

Perhaps the most promising is the least developed: wet ALife. If we build a system that most people agree on calling ‘alive’, most likely it will be from wet ALife. There are several challenges already mentioned, but there seems to be no inherent limit to building or finding alternative lifeforms, either artificial or extraterrestrial. It is a blind guess to try to say whether we will have detected or created life different from the one that evolved on Earth by 2050. Still, soft ALife and hard ALife seem to have inherent limits (derived from the limits of formal systems), so we might as well expect the most from wet ALife. In any case, it can certainly contribute to a ‘general biology’.

Physics can also contribute in this direction. Already, research in self-organization [319, 320] and active matter [321] has contributed to understanding the properties of living systems. Very likely, in the next few decades, advances in physics will enhance our perspectives on what we consider to be living, how it evolved, and where it might lead.

There is still no agreed-upon definition of life. Biological systems are perhaps too complex for a sharp definition. As we have seen, ALife can help to understand the general properties of living systems. This can benefit biology and engineering, gaining insights into life and being able to build artificial systems exhibiting properties of the living.

4.5 Toward a quantitative understanding of life

Raymond Goldstein[1] and Patricia Bassereau[2]
[1]University of Cambridge, Cambridge, UK
[2]Institut Curie, Paris, France

4.5.1 General overview

More than half a century ago, physics and biology came together to understand the DNA double-helix structure, one of the most important discoveries of the 20th century. At present, the complexity of biological structure and function and the observed emergent phenomena in biology delineates a well-defined research domain where physics meets biology. Emergent phenomena include processes where larger entities exhibit properties that their simpler constituent entities (and the naive superposition of their properties) do not exhibit and, as a consequence, something new emerges from collective behaviour that could not be predicted from its constituent parts. Complexity applies to key open questions in the physics of the many-body interacting systems, but it finds its most natural setting in biology, from the synchronized dynamical behaviour of the human brain or the beating of heart cells to the concerted action across multiple scale of biomolecules and cells in tissues and organs. The cascade of energy and information across the many biological levels defines a new field of research, the thermodynamics of information, which starts at the Maxwell demon paradox and the interpretation of the second law and may lead us far beyond.

4.5.2 Physics meets biology over the last 60 years

Physics and biology have been intertwined since the dawn of modern science (figure 4.15), from Antony van Leeuwenhoek's use in 1700 of advanced optics to reveal the hidden world of microscopic life [322], to Bonaventura Corti's (1774) discovery of the persistent fluid motion inside large eukaryotic cells [323], Robert Brown's 1828 study of random motion at the microscale [324], and Theodor Engelmann's determination in 1882 of the wavelength dependence of photosynthetic activity [325]. Although at the time it might have been difficult to define the precise disciplines of each of these scientists—perhaps they were all 'natural philosophers'—in hindsight we can see clearly the way in which their discoveries impacted both biology and physics. Despite this long history of discoveries at the boundary between the two fields, and the innumerable fundamental contributions to both disciplines over the long arc of time, the field of *biological physics* as a discipline within the research enterprise of physics has only risen to great prominence since the postwar era, particularly since the mid-1980s. We are now at the point that most academic physics departments have an identifiable group in biological physics alongside those in high energy, condensed matter, atomic and astrophysics. In this introductory section we will review some of key developments over the past 60 years in order to identify what we see as key intellectual threads that run through that history, to set the stage for the forward-looking sections that follow.

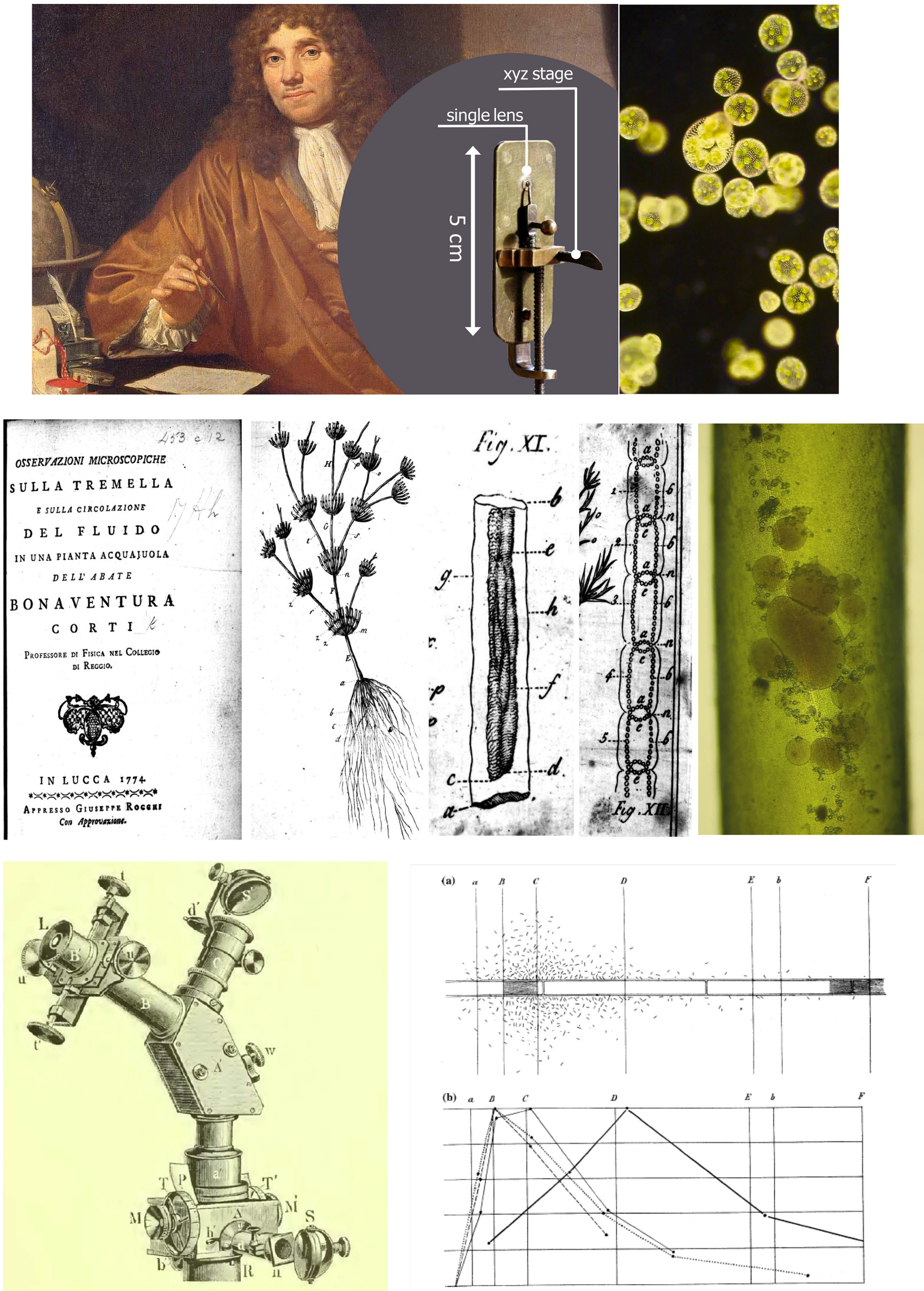

Figure 4.15. Historical connections between physics and biology. Top row: Antony vanLeeuwenhoek, his microscope, and the alga Volvox that he discovered. Middle: BonaventuraCorti's celebrated treatise on cytoplasmic streaming, with drawings of aquatic plants he studied,and a modern image of the plant Chara corallina. Bottom: The micrscope with which Theodor Engelmann's determined the action spectrum of photosynthesis by visualizing the accumulation of aerotactic bacteria along an alga illuminated by the solar spectrum. This image has been obtained by the authors from the Wikipedia website, where it is stated to have been released into the public domain. It is included within this article on that basis.

We begin by noting that each of the discoveries highlighted above was made possible by state-of-the-art *microscopy* that was able to reveal phenomena that had escaped previous notice. That technological advances often translate into scientific discoveries is a familiar pathway in science, but it takes more than just a new piece of kit to lead to true progress. Indeed, there is often a separate but equally important collective aspect of the scientific community at work, in defining the questions, bridging disciplines, and training students.

Our choice of 60 years for this overview was made to capture several of the most important immediately postwar advances at the interface of biology and physical sciences which, in an interesting historical twist, clustered around the same time (figure 4.16). These were the 1952 elucidation of the dynamics of action potentials in neurons by Hodgkin and Huxley [332] and the theoretical work by Turing in that same year showing that chemical reaction-diffusion systems can exhibit spatio-temporal patterns, followed in 1953 by the discovery of the structure of DNA by Franklin, Watson, and Crick [333, 334] using x-ray scattering methods. As in the discussion above, the work of Hodgkin and Huxley built very much on developments in experimental methods; in this case it was the invention of the 'voltage clamp' by Cole [335], a device that utilizes feedback to maintain the voltage across a membrane at a set value, that allowed the properties of ion channels in membranes to be studied as a function of (controlled) voltage. The idea of a voltage-dependent channel conductance was the key to understanding not only neuronal dynamics but 'excitable media' in general [336]. Experimental methods from physics were of course also key to the DNA structural work, which was an offshoot of the original development of x-ray scattering to determine crystal structures based on Bragg's law [337].

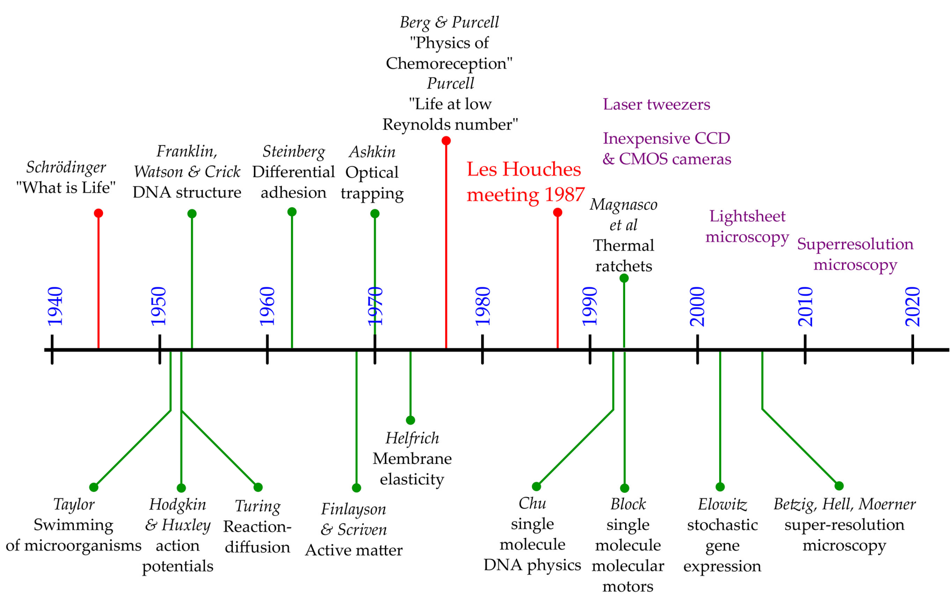

Figure 4.16. Timeline of significant developments in biological physics from the mid 20th century onwards.

Turing's work had little to do with technological advancements and his research would not generally be considered originally to be part of academic physics; it might be categorized instead as one of the earliest works in *mathematical biology* [338]. Using very simple mathematical models of two interacting chemical species, he showed that a system of chemical reactions that would, in the absence of diffusion, be linearly stable, could be rendered unstable by diffusion, leading to spatially periodic patterns. In this context, this mechanism was a revolutionary idea, although spatio-temporal symmetry breaking had already been understood in fluid mechanical contexts at least since the time of Rayleigh's work on thermal convection [339]. Yet, the relevance of the Turing mechanism to actual biological systems has been debated ever since [340], in part because in its standard form the mechanism requires the diffusivities of the two chemical species to differ by what seems to be an unphysical amount. It was physicists who provided the first experimental realizations of the Turing instability by reducing the diffusivity of one of the chemical species through binding to a gel substrate [341–343]. While it has also been unclear how to square highly regulated biological patterning with the concept of spontaneous symmetry breaking, it is important to emphasize that Turing's work introduced most of the concepts now used in the study of how biological form develops, the most important of which is that a system of reacting and diffusing chemicals can form spatio-temporal patterns [344]. It was within applied mathematics that reaction-diffusion dynamics were first applied to pattern formation in biological populations such as the spiral waves and chemotaxis exhibited by the slime mold *Dictyostelium discoideum* [345], but these ideas eventually resonated with physicists studying such pattern formation [346]. There is now ample evidence that these basic ideas are correct.

Like Turing's paper, Taylor's (1951) work [347] that offered the first explanation for the swimming of microscopic organisms in a viscous fluid sat not within physics proper, but rather in fluid dynamics (in the British school of applied mathematics). This and subsequent works by other applied mathematicians such as Lighthill [348] clarified the physics behind self-propulsion in the absence of inertia by exploiting the theoretical simplifications inherent in the dynamics of asymptotically slender filaments such as eukaryotic and prokaryotic flagella. These developments were later taken up by physicists and caught the attention of the community thanks to Purcell in his celebrated 1977 essay on 'Life at low Reynolds number' [349], and in his work with Berg [350] that highlighted physical considerations in the phenomenon of chemoreception. This is one of the many important examples in which scientists who came from more established areas within physics moved into biological physics, helping to legitimize it within the broader community.

These intellectual strands that were developing in the 1970s were very much apart from the molecular-biology-centered world of mainstream biology at the time, and indeed they were distinct from much of the field known as 'biophysics' that was focused on proteins, ion channels, and electrophysiology, but they would prove to be harbingers of the development of biological physics as a distinct discipline. Another excellent example of the biological research beyond the molecular level is that of Steinberg on the cellular organization in tissues. In his 1962 paper on this subject

[351], he formulated a differential adhesion hypothesis for the self-organization of cell types within tissues. Although the dominance of the molecular view of biology at the time meant this work gained little traction in the community, its enthusiastic uptake by physicists in the 1990s [352] led to an explosion of work on the subject of tissue biomechanics that continues to this day.

The gradual transformation of the field of solid-state physics into what we now term condensed matter physics began in the 1960s and early 1970s with the intense interest both in phase transitions and critical phenomena and also in the physics of liquid crystals. In moving away from the static description of precisely ordered solids, the field naturally began to focus on phenomena on length scales larger than molecular, and to utilize continuum descriptions such as those familiar from the Landau–Ginzburg theory for superconductivity, with particular emphasis on the formulation of scaling laws [353] in polymer physics, building on the foundational work of the theoretical chemist Flory [354]. An important strand of this research concerned the description of interfaces, as involved in phenomena such as wetting, and also in pattern formation during solidification [355]. The emphasis on continuum approaches was particularly significant in the field of liquid crystals, where the intellectual leadership of the French school led by De Gennes proved so important in creating a worldwide community of experimentalists and theorists. It was out of this field that came the seminal work of Helfrich in 1973 [356] on the elasticity of fluid membranes. The particularly simple form of this energy functional, dependent only on the shape of the membrane, allowed for an explosion of analytical and numerical work on a range of problems in membrane physics. With these advances, it was but a small step to move from the description of idealized membranes as found in lipid vesicles to real biological membranes with all their complexity.

On the experimental side, the arrival of affordable and relatively easy-to-build optical trapping setups in the early 1990s enabled quantitative experiments at the single-molecule level that were simply not possible previously. The first single-molecule experiments on the stretching of DNA in 1992 [357], combined with later theory [358], enabled the precise quantitative understanding of semiflexible polymers, opening up the understanding of aspects of chromosome structure and function in the cell. Likewise, pioneering work in 1993 on the stepping of motor proteins along biofilaments [359] revealed the stochastic nature of the motion of molecular motors and almost overnight placed their dynamics squarely within the field of non-equilibrium statistical physics. At roughly the same time, the paradigm of stochastic rectified motion on periodic potential energy landscapes (aka ‘Brownian ratchets’) received intense focus as the model for molecular motors [360–363] and also for the unidirectional motion exhibited by plants [364]. Together these helped to launch a subfield of stochastic thermodynamics, leading to principles for complex systems such as synthetic machines [365], including collective effects [366], with applications to such systems as hair cells in the ear [367]. From the mid-1990s onward, the advent of relatively inexpensive CCD and CMOS cameras and associated image processing techniques, including particle tracking methods [368], meant that the barrier to experimental work in this area was dramatically lowered.

Later developments in the new millennium saw a whole new generation of affordable high-speed cameras, further enabling the worldwide study of fast phenomena such as flagellar synchronization [369] and cell motility [370]. Finally, the development of soft lithography for microfluidics [371] further broadened the experimental base for studies at the microscale.

From all of the above, we see that fields that in the postwar period were considered apart from the core of physics, including fluid mechanics, transport theory, and much of continuum physics), were embraced by the biological physics community as its focus turned toward phenomena on scales from nanometers to millimeters, where so many cellular phenomena occur. It was perhaps only natural that physicists in this new era would reject [372] the view [373] that their role in biological research was simply to provide instrumentation or better intermolecular potentials in the service of the questions biologists had framed. Rather, the field has prospered precisely because physicists and biologists have joined together to pose the questions that guide the field. Finally, as the ranks of academia working on biological physics grew throughout the 1990s to the point that a new generation of PhD students was trained and themselves moved into academia, the whole field gained critical mass to the point of being among the fastest growing divisions within national physics societies.

4.5.3 Challenges and opportunities for the future

In this section we speculate on the future directions of a few of the many areas within biological physics, aiming for an overview rather than an encyclopedic account.

4.5.3.1 Soft matter and self-organization: from active matter to cell and tissues

In a remarkably insightful article in 1969 [374], Finlayson and Scriven considered the possibility of various types of hydrodynamic instabilities arising from what they termed 'active stresses'. These are contributions to the stress tensor arising from gradients of scalar fields (e.g., the concentration of some solute) that represent the conversion of chemical energy to kinetic energy, typically manifest by a pattern-forming instability. This is the essence of what we now term 'active matter'; systems in which there is injection of energy at small scales that display coherent structures on scales large compared to the microscopic constituents [375–378]. Early examples of this notion are found in new physical models for membranes that include the non-equilibrium activity of proteins that transfer ions across membranes via external sources of energy (light, ATP hydrolysis, electric fields) [379, 380]. Other examples include collections of molecular motors and microtubules [381], where the consumption of ATP by the motors powers translocation that leads to filament motion and self-organization, and in collections of self-propelled particles, in theory [382–384] and experiment [385], where coherent structures arise from hydrodynamic interactions between the organisms (figure 4.17).

These general concepts and tools were used in parallel to study the organization of living systems and of engineered ones. On the biological side, 'active membranes'

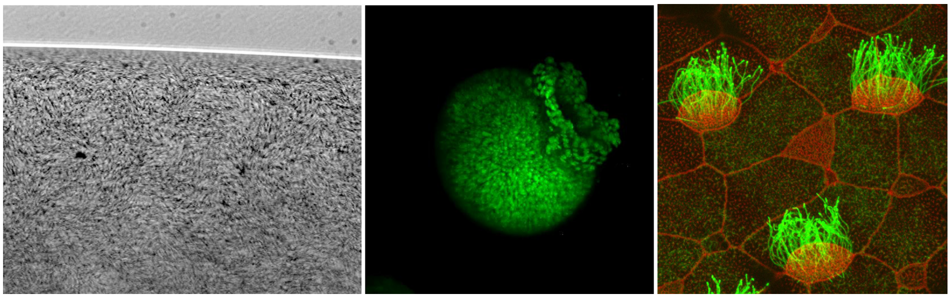

Figure 4.17. Active matter on multiple scales, from individual cells to confluent tissues. From left to right: bacterial 'turbulence' in a suspension of B. subtilis (reproduced from [326], CC BY 4.0), embryonic inversionin Volvox (reproduced from [327] CC BY 4.0), and a tissue with multiciliated cells (Mitchell Lab, Northwestern University).

have been generalized to membranes with a cortex and in contact with polymerizing actin filaments [386, 387]; they contributed to elucidate the non-equilibrium origin of the flickering of the red blood cells [388], although a consensus is still not reached. They have been applied to active gels [389, 390] made of polar dynamical cytoskeleton filaments and active crosslinkers like molecular motors that generate flows and stresses, features that do not exist in passive gels. Active gels have been instrumental for understanding actin flows in different cellular features: cell motility, blebbing, division, etc [390]. Liquid crystals principles have been extended to active nematics, smectics, or cholesterics. The power of this continuous description is that the model of active nematics can describe as well the self-organization of patterns and the flows for actively moving entities such as bacteria in films, anisotropic motile synthetic particles, animals flocks, or that of cells in living tissues during development, embryos [391]. Of course, non-equilibrium activity also affects phase separation [392]. A rapidly growing body of work has shown that liquid–liquid phase separation of multivalent assemblies occurs in the cytoplasm of cells, forming membraneless compartments [393], or in the nucleus [394]. However, Oswald ripening and thus droplet size is probably limited by active enzymatic reactions [395].

While many of the ideas and experiments on active matter systems implicitly use as a reference dilute suspensions of the motile entities, in recent years there has been a gradual shift of focus toward "confluent' tissues, typically quasi-two-dimensional sheets whose constituent cells are in space-filling contact with their neighbors. This leads to a wholly different class of phenomena and theoretical models that touch on some of the most significant issues in developmental biology. During development, tissues like epithelia are very dynamic since cells continuously die or divide and in addition, active forces are exerted at the cell–cell junctions. Thus, in spite of their morphological similarity with foams, epithelia differ strongly from them [396, 397], and non-equilibrium principles are necessary to describe their homeostasis [398, 399]. In addition, many tissues bend, fold, and even invert their topology during embryogenesis [327]. Through a combination of advances in experimental methods such as light-sheet imaging [400] and use of 'organoids' [401] to study the early stages of multicellularity, and theoretical models addressing geometric rearrangements of tissues [402], we anticipate that the near future will see significant progress in understanding many key issues in the biomechanics of development and differentiation.

So far, very little work has been done to integrate at the cell scale all active exchanges that occur between compartments, with the plasma membrane and with other cells. At a single compartment level such as the Golgi apparatus, they directly affect its shape [403]. Apart from during division, cells have to maintain their shape, area, and volume in spite of these multiple fluxes, and how they manage is a recurring issue in biology. Modeling these exchanges and understanding how homeostasis at the cellular scale is achieved remains one of the future challenges for physicists and cell biologists. The following section addresses how homeostasis is maintained in tissues.

4.5.3.2 Deciphering the physical principles of mechano-chemical networks that control homeostasis, shape, and size of living entities

In his book *On Growth and Form* [404], D'Arcy Wentworth Thompson raised the question of how physical forces contribute to determine the size and shape of living organisms and thus initiated the field of 'mechanobiology'. In addition to the non-equilibrium principles that govern cellular assemblies, mechanics also plays a key role. Indeed, cells exert, sense, and respond to external forces. For instance, the spreading velocity of cellular migration depends on the stiffness of the underlying substrate [405]. Cells exert forces on their environment generally using dynamical actin networks and the contractile actomyosin machinery. To a large extent, cell mechanosensitivity depends on proteins embedded in the plasma membranes that are linked on the extracellular side to specific ligands of the external matrix or to similar membrane proteins of a neighbouring cell in tissues. On the intracellular side, they have cryptic binding sites that unfold and allow connection to the actin cytoskeleton in a load-dependent manner. Actin structures are themselves mechanosensitive [406]. Moreover, mechanical forces trigger biochemical response and signalling pathways (i.e., cascades of biochemical reactions with positive and negative feedback loops). A revealing illustration is provided by stem cells (i.e., non-differentiated, pluripotent cells) that differentiate into very different cell types—neurons, muscle, or bones—depending on the stiffness of their micro-environment [407]. Mechanical cues are thus transduced into biochemical signals, and integrated with genetic and chemical signals to modulate diverse physiological processes. In addition, there is constant cross-talk between biochemistry and mechanics during mechanotransduction, which is itself a part of the early development since developmental genes can be switched on by internal stresses accumulated during the growth of the embryo [408]; it is also involved in cancer development [409]. Physics and bioengineering have strongly contributed to this field, in particular by developing many tools for measuring forces at all scales, from the single mechanosensitive molecules to stresses in tissues [410]. On the biochemical side, complex signalling networks have been identified in cell-extracellular matrix adhesion mediated by integrins [411, 412], or adherens-junction in cell–cell contacts [413]. Mechanosensitive channels, in particular ‘piezo’ channels, the main type of molecular force sensor in eukaryotes [414], are present in cell membranes; they let ions flow when they mechanically activated, which also triggers a cascade of biochemical signals, but have been less studied.

Cross-talk exists between these different signalling pathways that are all integrated at the cell level. Systems biology approaches are certainly essential in understanding these complex regulatory networks and how cells manage their mechanical interactions with their environment. But models based only on gene ON/OFF circuits are insufficient to understand how tissue integrity is preserved against mechanical stresses, extensile, or compressive, and during rearrangements of cells, and how tension homeostasis is set. The existence of cellular rearrangements in tissues implies the remodeling, destruction, or creation of the connecting structures between cells in a coordinated fashion (figure 4.18). In some rare cases, the feedback loops between mechanics and signaling are known [415]. Since ever-more force sensors are available to quantify stresses across scales, these measurements have to be integrated with the networks in cells to develop models that explain the emergence of larger-scale behavior from the interactions of their molecular components inside cells. Thus, it appears that it will still be some time before we have a final answer to D’Arcy Thompson’s questions: what limits the growth and division process, and what determines the size and shape of organs or animals?

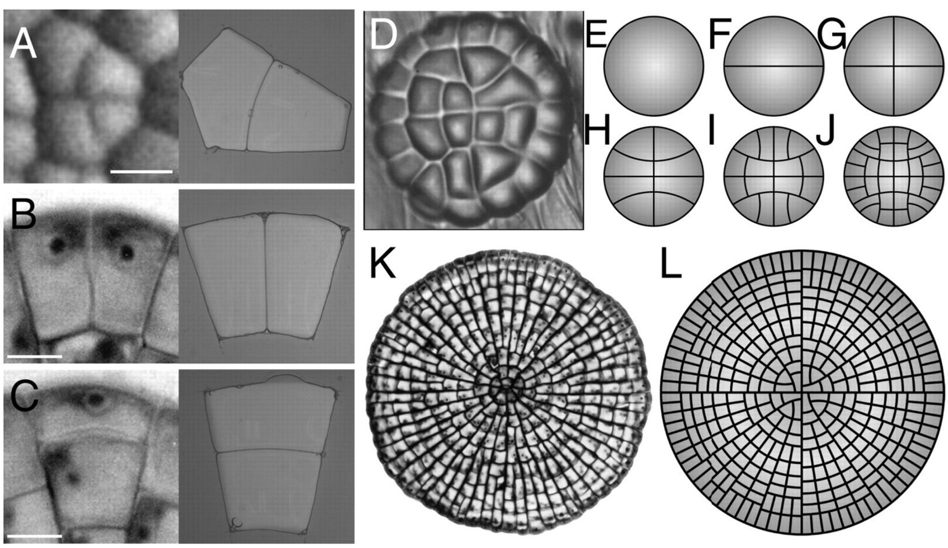

Figure 4.18. Cell division in the plant cells and similarity to shapes of soap bubble. Reproduced from [328] CC BY 4.0. (A–C) Patterns of cell division in various plants (left) compared to shapes of soap bubbles (right). (D) A large-scale geometry compared to mathematical results (E–J) in which a rule derived from soap bubble physics is iterated for uniform growth in the region. (K, L) as in (D–J), but for marginal growth.

4.5.3.3 Some perspectives on brain functions, for soft matter and computational neuroscience

Neuroscience has a singular place in biology. It is at the cross-road between disciplines, since the brain is not only an organ but controls locomotion, sensing, memory, decision making, and, at least in humans, feelings, consciousness, etc. The brain is a complex, temporally, and spatially multiscale structure. From the perspective of physics, it can be of course investigated at the cellular level. But since neurons communicate and form circuits and interconnect functional areas in the brain, it is better described as a hierarchical network (termed the 'human connectome' [416]). In addition, neurons are embedded in glial cells that protect them, but also contribute to some signaling functions. Two of the most challenging goals in science continue to be on the one hand the generation of the complete map of the neural connections in a brain and, on the other, to understand from a molecular point of view how signals are produced and transmitted and how the network builds up, in order to decipher how this incredibly complex structure can produce complex cognitive functions. The breadth of challenges for physicists of the future is too wide to list in totality; here we mention two of them, on axonal signaling and for computational neuroscience.

Neuronal cells have unique properties—mechanical, geometrical, and internal organization—and are actively studied *per se*. Strikingly, not much has been done since the work of Hodgkin and Huxley [332] to revisit their model of axonal transmission based on electrical signals. In the standard action potential model, signal propagation is achieved by the voltage-dependent opening and closing of ion channels, largely ignoring the specific physical properties of cell membranes. Only T Heimburg has challenged it, by suggesting that a lipid phase transition occurs in the membrane in the course of the action potential, increasing membrane conductance [417]. However, no study has included yet the non-equilibrium effect due to the activity of the channels we discussed above. Thus, a comprehensive model of the propagation of the electric axonal signal is still missing.

Various methods are routinely used to image whole-brain activity and detect dysfunction and disease, including x-rays (CT scan), radioactivity (PET scan), and NMR (MRI). However, to image the neuronal networks with better resolution, neurosciences have greatly benefited from the most advanced developments in microscopy and imaging: methods for imaging in diffusing media allow deep imaging in the brain, light-sheet microscopy for volumetric imaging, and optogenetics to control neuronal circuits [418]. Model animals with small brain volume or that are moderately transparent (*Drosophila*, zebrafish) have also facilitated these studies. It is now possible to follow neural algorithms in living and freely moving animals as they vary their behavior [419]. Using whole-brain functional imaging, the brain of a zebrafish can be imaged during the different stages of the decision-making process about its swimming direction [420]. Betzig and collaborators achieved the *tour de force* of imaging the whole brain of a fruit fly with molecular contrast and nanoscale resolution using combined light-sheet and expansion microscopy [421]. It is thus now possible to locate individual neurons, trace connections between them, and visualize organelles inside neurons, over large volumes of brain tissue in 3D, albeit on fixed brains. These are only a few examples showing how the field is developing with the blooming of new optical techniques, pushing the frontiers of the observations. One obvious consequence is that with these advances will come enormous quantities of data, as in many other areas of cell biology in which volumetric imaging is used. One strategy to manage and analyse such 'big data' is obviously to use artificial intelligence and deep learning methods [422] to extract meaningful information. In addition, there is now a timely opportunity for computational sciences to develop approaches based on network science to provide integrated models of interactions in neurobiological systems. In fact, a new field termed 'network neuroscience' is growing, bridging network theory and experiments [423]. One might hope that creative developments in computational sciences in theory and in functional imaging will eventually allow us to unlock the neuronal code.

4.5.3.4 Emergence of life and physics of biological evolution: from the second law of thermodynamics to the selection of structures

In his 1944 book *What Is Life? The Physical Aspect of the Living Cell* [424], Erwin Schrödinger stressed the apparent paradox behind life: how can living organisms maintain an organized state and grow complex structures without violating the second law of thermodynamics that predicts an evolution towards maximized entropy? He resolved it by pointing out that Earth is not an isolated system, but receives energy from the Sun, and that living systems absorb energy. Moreover, he also used thermodynamic arguments to explain why an internal organizing factor that carries information (that eventually turned out to be DNA) is necessary for living systems to develop in an organized manner and replicate faithfully.

Likewise, the emergence of life on Earth cannot be explained by the second law, but rather (in part) by far-from-equilibrium thermodynamics and the concept of dissipative structures highlighted by Prigogine and others. While this issue is covered elsewhere in this volume, we note that there is a strong school of research, by no means universally accepted, supporting the idea that the evolution of life began from 'soup' of RNA molecules before DNA appeared. Yet, the mechanisms by which the nucleotide bases and sugars could be formed beforehand by prebiotic reactions are still not elucidated. There are promising attempts to produce artificial cells [425] and a significant body of work on life-inspired and out-of-equilibrium systems at the nanoscale [426]. Compartmentalization and the appearance of membranes are key steps during evolution. Recent work on active membraneless droplets suggested that they could have formed the protocells from which cell membranes could have appeared [427]. Bottom-up reconstitution of a synthetic cell with well-characterized functional molecular entities in vesicles can also help to understand the origin of life [428]. Conversely, with a top-down approach, the Craig Venter Institute has shown it is possible to recreate artificially genomes and minimal cells [429]. A synthetic minimal organism has been reproduced *in silico* by reconstruction of a complete set of chemical reactions [430]. Hydrodynamic models have also been used to understand how DNA may have replicated in early times: laminar thermal convection, present in submarine hydrothermal vents, can very efficiently accelerate the DNA replicating polymerase chain reaction (PCR) [431], which is enhanced when the molecules are trapped in porous rocks [432]. We expect these *in vitro* approaches will become ever-more important in the future. On the modeling side, considering the complexity of this interdisciplinary problem, there is a clear need for further development of related aspects of non-equilibrium physics.

4.5.3.5 Evolution of biological complexity

High on the list of fundamental problems in biology, just behind the origin of life and the nature of consciousness, is the origin of multicellularity. While the simplest organisms to appear on earth were no doubt unicellular, eventually life evolved to become larger, in the sense of having more cells, and also more complex, dividing up life's processes into ever-more specialized cell types [433]. It has been recognized since the time of Weismann [434] in the late 19th century that a great challenge is to understand the driving forces behind the transition to multicellularity, and, as pointed out by Huxley some years later [435], to identify the biological entities on which evolution acts [435]. While there can be obvious advantages to larger and more complex organisms, such as greater motility, avoidance of predators, and larger uptake rates of nutrients, there are also metabolic costs associated with the regulatory networks that control the organism and the cellular scaffolding that holds it together [436]. Recent work has begun to address these issues using green algae [437] and choanoflagellates (figure 4.19), the closest uni- and multicellular relatives of animals [438].

The ability to track *single* living objects (bacteria, yeast, or cells) using microfluidic devices and to analyze their lineage during multiple rounds of cell division will open the way for new discoveries when these experiments will be coupled to external perturbations. Among the most promising approaches to understanding the origins of biological complexity involve the use of artificial methods to put evolutionary pressure on extant organisms. A prime example of this is recent work on yeast, in which repeated rounds of centrifugation of growing cultures, selection of

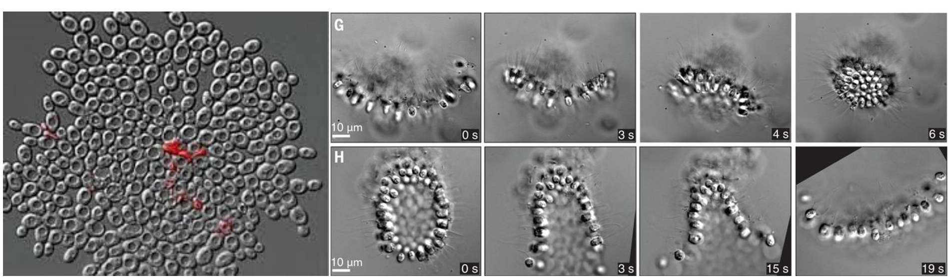

Figure 4.19. Novel multicellular organisms. Left: 'snowflake yeast'. Reproduced from [329] CC BY 4.0 formed via repeated rounds of selection for faster settling speed under centrifugation. Right: snapshots during the curvature inversion of a sheet of choanoflagellate cells comprising the organism C. flexa, as triggered by light. Reproduced from [330] CC BY 4.0.

the fastest sedimenting fraction (containing the largest organisms), and subculturing of that fraction produces 'snowflake yeast' (figure 4.19), a genuine multicellular variant [329]. This 'experimental evolution of multicellularity' enables a whole range of questions in the origins of multicellularity to be addressed, and we anticipate significant developments in this area in the coming years. It is clear that ideas from statistical physics [439, 440] will be important for the analysis of the data arising from these studies. In sum, we can see the emergence of a field centered around the physics of biological evolution, using concepts from condensed matter such as frustrated states and glasses to describe transitions during evolution [441]. With the appearance of ever-more data from advances in experimental methods due to technical developments, we can imagine that more theoretical models will be designed to understand how living systems cope with the second law to adapt to their environment.

4.5.4 The future

While predictions are always difficult, especially about the future, we offer a few final words on the greatest challenges in the field of biological physics may address in coming years. Clearly, the most fundamental, unsolved problem is the origin of life. As we have touched on here and as discussed in greater detail elsewhere in this volume, a range of highly interdisciplinary efforts appears poised to make significant progress on this problem. The issues concern the bootstrapping problem of how a truly self-replicating system can arise biochemically, and also the geophysical conditions that are amenable to such a development. Likewise, the nature of consciousness remains mysterious, but we can anticipate that the continued development of probes of neuronal structure and organization will point the way toward a deep understanding of this emergent phenomenon. In the realm of developmental biology, it is clear that the rapid explosion of experimental methods to probe cellular fate determination and global regulation combined with physical concepts regarding spatio-temporal patterning will continue apace, and we can look forward to a deeper understanding of the *regulation* of development. Even such issues as the regulation of limb size are, at this point in time, not resolved; their study in model organisms will continue to be an important research endeavor. Ultimately, we may hope that these physical methods will contribute to understanding the origins and control of the unregulated cell division that is at the heart of cancer. At the subcellular scale we still lack a comprehensive understanding of such complex machines such as the ribosome, which work in confined environments or with a limited energy supply. At the more macroscopic scale, the use of *in vitro* evolution methods will surely continue, providing a platform for the true quantitative understanding of evolution in the natural world. Coupled closely to this will be an increasing focus on what might be termed 'physical ecology', the study of communities of organisms coexisting with their natural habitat (figure 4.20). As of this writing, the world is wrestling with a global pandemic and it seems natural to expect much future research on the interplay between viruses and their hosts.

Figure 4.20. Two aspects of physical ecology. Left: marine algal blooms (green) in the Baltic Sea as seen from a European Space Agency satellite (reprinted from [331], courtesy of ESA). Right: the blue glow of bioluminescence triggered in breaking waves at a beach (photo courtesy of Gergo Rugli).

Acknowledgements

The work of P Bassereau's research group is supported by the FRM (Fondation pour la Recherche Médicale), the ANR (Agence Nationale pour la Recherche), the AFM (Association Française pour les Myopathies), the CNRS, and Institut Curie. R E Goldstein acknowledges support from the Engineering and Physical Sciences Research Council, through Established Career Fellowship EP/M017982/1, Investigator Award 207510/Z/17/Z from the Wellcome Trust, Grant 7523 from the Marine Microbiology Initiative of the Gordon and Betty Moore Foundation, and the Schlumberger Fund.

4.6 The emergence of life: the Sun–Earth connection

Sarah Matthews[1] and Frances Westall[2]
[1]Mullard Space Science Laboratory, University College London, London, UK
[2]Centre de Biophysique Moléculaire, Université d'Orléans, CNRS, Orléans, France

4.6.1 Introduction

Life on Earth is sustained by radiative emission from the Sun, as well as by heat emanating from the outer liquid core and the mantle. Indeed, it appears that this energy may have been essential for the very emergence of life on Earth. The Sun's emission varies on all timescales at which it has been observed, as a consequence of the generation, emergence, and evolution of its intrinsic magnetic field. There is increasing evidence that solar radiative changes have had an influence on Earth's climate through different mechanisms. In addition to variable radiative emission, the Sun also emits energetic particles, the flux of which changes in time depending on solar magnetic activity. Variations of the solar particle emission cause changes in the ionization of the Earth's atmosphere, its global electric circuit, and the ion-induced nucleation and condensation nuclei in the Earth's atmosphere. Understanding climate and the emergence of life on Earth thus requires knowledge of the natural variations of the radiative and particle fluxes received from the Sun. The main scientific questions are:

- What is the influence of solar radiation on the origin of life on Earth and possibly elsewhere in the Solar System and the Universe?
- How does the solar emission vary at different timescales, from seconds to centuries and millennia?
- What is the spectral dependence of the solar radiative variability?
- How does the magnetic field affect the radiative and particle emission of the Sun?
- How does solar variability affect the climate and life on Earth?

4.6.2 Challenges and opportunities

A major challenge is to get a better understanding of the environmental conditions reigning on the Earth when life emerged, including the physical characteristics of the Sun and their influence of Earth's early environment. Future missions to search for life on icy satellites, such as the NASA Europa Clipper flyby mission (2024) and the ESA JUICE flyby mission (2023) to Jupiter's moon Europa, and proposed missions to Saturn's moon Enceladus, will hopefully provide us with some indications as to whether life emerged on these satellites, vital information to understand the origin of life in general, and the exact role of solar radiation in prebiotic processes. It is now generally accepted that the recent climate change on Earth has been influenced by human activity. It is also clear that continued emissions of greenhouse gasses in the atmosphere will cause further warming and changes in all components of our climate system. Therefore, improving our knowledge of the solar radiative and particle emission and of the mechanisms behind their changes is fundamental to advance our understanding of the climate system on Earth as well as to improve the accuracy of models predicting future climate scenarios.

4.6.2.1 The influence of radiation on environmental conditions and habitability of the early Earth

Environmental conditions and therefore habitable conditions on the early Earth were dependent on a complex interplay between the early evolution of the Sun and the surface volatile envelope of the early Earth. Solar evolution must have played an important role, controlling atmospheric chemistry and temperatures and therefore the existence and state of water at the surface. It also provided the energy for certain critical prebiotic molecular reactions and may even have destroyed some molecular species. Pre-ozone levels of UV flux to the Earth's surface certainly affected processes and early metabolisms. The evolution of the Sun throughout the early period of geological history when life was getting a foothold therefore contributed to the early evolution of life. Thus, the emergence and early evolution of life must have depended a lot on the Sun and its physical environment.

4.6.2.2 Habitability

It is useful here to begin with a brief discussion of the concept of habitability and the Habitable Zone (HZ) around a star (see figure 4.21). Originally defined as the zone in which liquid water exists at the surface of a planet [442], it is now understood that other parameters, such as cloud cover, need to be taken into account and that the habitable zone can be extended. Indeed, with respect to icy satellites and exoplanets, the concept changes. Here, bodies of liquid water below icy crusts are maintained by internal planetary processes and tidal heating due to the gravitational resonance of the icy satellite with the main planet. These bodies are far outside the traditional HZ. The timing of, and the processes leading to the emergence of life on Earth were

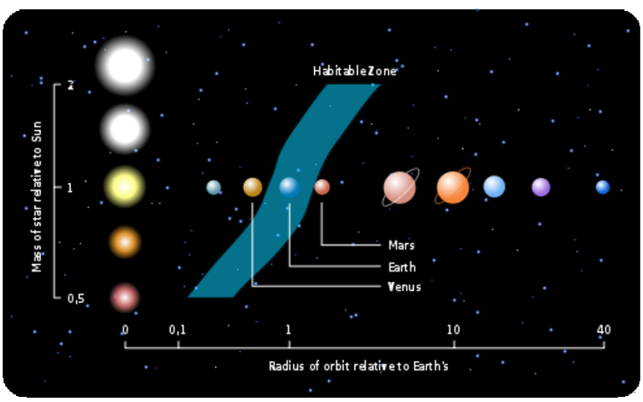

Figure 4.21. Classical view of the Habitable Zone predicated on the existence of liquid water at the surface of a planet. The planets of the Solar System, with the exception of Earth, are either too hot or too cold. This image has been obtained by the authors from the Wikipedia website, where it is stated to have been released into the public domain. It is included within this article on that basis.

predicated on the initial establishment of habitable conditions, namely the presence of liquid water at relatively low temperatures less than 120 °C, the presence of prebiotic molecules, and the availability of energy sources. However, prior to the installation of these initial conditions, the Earth had already undergone a rather violent history. Understanding when exactly the Earth became habitable is challenging because we are lacking a well-preserved rock record covering the first billion year's history of the planet: The earliest crust was thoroughly recycled through tectonic processes and impacts during the Hadean (4.56–4.0 Ga, 1 billion years ago = 1 giga annum = 1 Ga) and the Eo-Palaeoarchaean (4.0–3.2 Ga). It is thus necessary to base our understanding on theoretical considerations coming from modeling, comparative planetology, and the use of any information on the oldest, well-preserved rocks on Earth. They date back to 3.5 Ga, i.e., one billion years after the consolidation of the planet.

4.6.2.3 Early atmosphere(s) and oceans

Life arose within the ephemeral envelope of liquid and gaseous volatiles that enclosed a solid crust and mantle of silicate rocks, which in turn enclosed a core of solid and liquid iron, nickel, sulphur, and other elements, all fractionated out during the formation of the planet. The Earth consolidated from planetesimals at about 4.56 Ga. Initially heated by short-lived radio nucleotides, such as ^{28}Al, the early magma ball slowly cooled down, degassing a first, mainly H_2 atmosphere [443]. At this stage, too hot and without liquid water, the Earth was uninhabitable. One or more Moon-forming impacts reduced the Earth again to a magma ocean, either totally or at least partially, the last of which occurred between 4.9 and 4.43 Ga [444] or about 4.45 Ga according to [445]. The magma ocean rapidly crystallized out within at most a million years (see, e.g. [446]), outgassing a short-lived steam atmosphere of H_2O and CO_2 originating from an oxygenated mantle, the CO_2 outgassing earlier because of its lower solubility in magma, and H_2O later because of its higher solubility [447]. Oxygenation of the early mantle, as deduced from chondritic D/H ratios, was critical to the composition of the outgassed atmosphere. A reduced mantle (i.e., low in oxygen) would have produced an atmosphere rich in atomic hydrogen that would have been more rapidly lost to space due to interactions with the solar wind and the radiation conditions [448], whereas a CO_2-rich atmosphere would result in IR cooling of the upper atmosphere, thus decreasing atmospheric expansion and the risk of erosion by solar processes [447]. It is estimated that the early atmosphere was dense, comprising about 100 bar CO_2 and a couple of hundred bars of H_2O [449], the latter raining out as ocean.

The contribution of volatiles to the Earth's envelope is the subject of much debate. Different hypotheses oscillate between the accretion of the planet from volatile-rich planetesimals and their degassing with minimal importation from later accreted materials [450], to major importation through carbonaceous chondrites, micrometeorites, and interstellar dust particles (IDPs) [451]. The surface of the crust would have needed to be sufficiently cool for the volatiles to condense out into a liquid ocean. Models of the cooling of the crust correlated to the gradual waning of the flux of large impactors that boiled off portions of the early oceans suggest that habitable conditions, i.e., temperatures below 120 °C, could have been present as early as 4.4 Ga [452].

The composition and density of the early atmosphere were crucial to both environmental conditions and the existence of an early ocean. Both were partly influenced by the evolution of the young Sun. Extreme ultraviolet (EUV and XUV) radiation (1–103 nm) was the main source of energy and ionisation in the upper atmosphere, thus influencing atmospheric escape [453]. In particular, the rotation of the young Sun must have affected the flux of XUV coming to the Earth. Lammer *et al* [454] estimate that this flux could have been about 15 times greater than at present for a slowly rotating young Sun, and up to 150 times greater for a fastly rotating young Sun. Both XUV and EUV radiation from the young Sun were therefore higher during the Hadean (4.56–4.0 Ga) but decreased with time, together with x-ray and magnetic activity [455]. However, in what is known as the Faint Young Sun Paradox, the lower luminosity of the young Sun (about 70% lower than today) at the time that it became a main-sequence star meant that during the Hadean/Palaeoarchaean period the Earth must have been at the outer, cold edge of the HZ around the Sun, with frozen water. Nevertheless, evidence exists that this may not have been the case and there are numerous hypotheses concerning the temperature of the early atmosphere, depending on the composition and partial pressures used in the models. While one model indeed suggests that water at the surface of the Earth should have been frozen [456], others suggest various mechanisms for warming up the atmosphere to get a liquid water surface. For example, surface temperatures on the Earth were and still are influenced by absorption of visible and near infrared radiation from the Sun. Absorption of the internal infrared radiation warms the atmosphere and some of this radiation is reflected back down onto the Earth's surface, thus warming it up and contributing to the total radiation energy received from the Sun. On the other hand, the Earth is cooled by emission of thermal infrared radiation, akin to black body radiation [457]. These processes are in thermal equilibrium.

Both CO_2 and H_2O, the so-called greenhouse gases contributing to warming up of the surface of the Earth, are believed to be present in the early atmosphere of the Earth. H_2O is a stronger greenhouse gas than CO_2, absorbing infrared radiation over a wide range of wavelengths [458]. Other greenhouse gases that may have been present include CH_4, which warms the atmosphere far more efficiently than does CO_2. While today most Methane in the atmosphere is of biogenic origin, abiotic production of Methane through serpentinisation reactions of the mafic crust, releasing a small but significant fraction of CH_2 into the atmosphere [459], could have been predominant on the Hadean Earth. The only way CH_4 can be removed from an oxygen-poor atmosphere is through photolysis at wavelengths below 145 nm, thus producing CH_3, CH_2, and CH radicals [460]. This gives an estimated lifetime between 10 000 and 20 000 years for CH_4, much longer than the lifetime of CH_4 in today's oxygenated atmosphere, equal to only 10 years. There is a caveat to high abundances of CH_4 in the atmosphere. If they are too high an organic haze is formed [461], which has a significant albedo effect, reflecting the Sun's radiation back into space and thus cooling the surface.

This albedo, expressing the fraction of solar radiation reflected back into space by the clouds and the surface plays a crucial role in the warming of the Earth. It is determined by the interaction between the variables mentioned earlier and governs the negative feedback of increasing infrared outbound radiation due to increasing temperature that leads to cooler surface temperatures [457]. Moreover, a recent study [462] has added the intriguing hypothesis that tidal heating between the young Moon and the Earth could have contributed to potential heat sources other than the faint young Sun. Ocean condensation from a primordial steam atmosphere may already have occurred around 4.4 Ga [452], but it is difficult to estimate exactly when temperatures became conducive enough to initiate prebiotic processes leading to the emergence of life.

Based on geochemical arguments, the volume of the early oceans is estimated to be between 30% and 50% larger than today, slowly losing mass through H_2O dissociation and hydrogen escape [463, 464]. The larger volume of water probably implied that the early protocontinents were largely submerged. Ocean salinity was likely similar or slightly higher (30%) than modern values [449, 465].

4.6.2.4 Prebiotic chemistry and the emergence of life

The basic initial conditions for prebiotic reactions to occur are the presence of liquid water, organic molecules, and a source (or sources) of energy (figure 4.22). Various catalyzers and other controls must also have been necessary to ensure that the kinetics of the reactions were fast enough and that the reactions went forward, i.e., were irreversible. Kinetic barriers are a requirement to ensure the irreversibility of these reactions. Robert Pascal and Addy Pross have been influential in highlighting the absolute need for such kinetic barriers, and the role of especially high-energy radiation in providing the initial free energy needed to set in motion the prebiotic processes leading to life [466, 467]. Since more than 150 kJ mol^{-1} is necessary to fuel

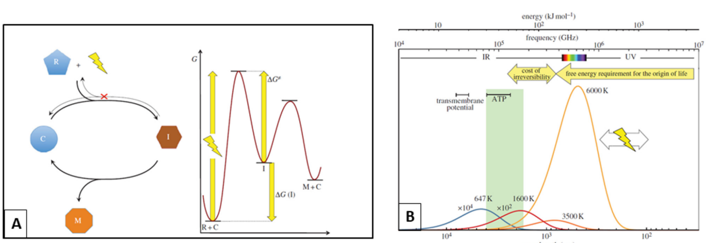

Figure 4.22. Kinetic energy barriers and sources of energy for the initiation of prebiotic reactions. (A) (left) Driving a catalytic cycle (R, reactant; C, catalyst; I, intermediate; M, downstream metabolite) to proceed unidirectionally by coupling with an energy source. (right) Irreversibility requires the waste of an amount of free energy corresponding to the kinetic barrier of the reverse reaction ($\Delta G^{\neq}$). (B) Comparison of different sources of energy available in planetary environments: electromagnetic radiation (correspondence with frequency and wavelength in abscissa), thermal energy (black body radiation curves displaying spectral radiance in ordinate: at 647 K, the critical point of water, blue line; 1600 K, representing typical Hadean magma temperatures red line; and 3500 or 6000 K, dark and light orange lines, surface temperatures of examples of M-stars or G-stars as the Sun, respectively) and lightning ($T \geqslant 10^4$ K). [467]

initial reactions, these authors point out that such high energy could not have been provided by the geothermal sources traditionally proposed as the source of energy for prebiotic reactions. Hadean lavas at 1600 °C would have produced only 70 kJ mol^{-1}. On the other hand, three other processes exist that could produce enough energy: (1) thermal energy at ~3600 Kelvin produced by shock metamorphism caused by impacts; (2) lightning in the atmosphere (a bolt of lightning can reach 29 725 °C); and (3) physical energy produced by photochemical processes at wavelengths of about 800 nm (150 kJ mol^{-1}). The latter is the dominating wavelength of radiation emanating from main-sequence G stars. The amount of 150 kJ mol^{-1} is the minimal energy needed to initiate self-organization of molecules in a manner that could not be reversed.

Photochemical dissociation of CO_2 and CH_4 in the atmosphere fueled by UV and lightning can yield organic molecules, such as HCN and CH_2O [468]. Strecker's synthesis of organic molecules in an experiment reproducing an early, reduced atmosphere of H_2S, CH_4, NH_3 and CO_2 subjected to lightning (spark discharge) was the first major prebiotic experiment undertaken [34]. Smallprim organic molecules were also formed through Fischer–Tropsch processes in the crust during the alteration of the ultramafic (rich in Fe and Mg) crustal rocks by through-flowing hydrothermal fluids that formed serpentine. This produced Methane, longer-chain alkanes, and more complex organic molecules [469, 470]. It has been suggested that organic matter may also have been degassed from the planetesimals forming the Earth [471, 472]. Nevertheless, the most important flux of organic matter to the early Earth must have come from space in the form of the carbonaceous meteorites (carbonaceous chondrites, CC), micrometeorites, and interplanetary dust particles (IDPs) that also contributed volatiles to the Earth's atmosphere and oceans [450, 473]. These materials can contain a huge variety of molecular compounds; for example, reference [474] estimated that more than 14 000 compounds exist in the Murchison meteorite.

Layers of impact spherulites from the early terranes in South Africa and in the Pilbara in Australia (3.5–3.2 Ga) confirm the continuing and significant flux of impactors during the Palaeoarchaean era (3.6–3.2 Ga, [475, 476]), but it is only recently that evidence for the presence of extraterrestrial carbonaceous matter found in these sediments has been documented [477]. This observation is remarkable because it is the first time that extraterrestrial carbonaceous matter has been found in terrestrial sediments and also the oldest evidence (3.33 Ga). Interestingly, micrometeorites have been proposed to be the carriers of this ancient extraterrestrial carbonaceous matter.

The role of high-energy impacts has been invoked by reference [478] as an important source of energy with respect to the possible emergence of life on other planets, such as Mars and even exoplanets. The most important supply of the required high energies came from solar radiation, and this has important implications for the kinds of environments in which life could have emerged, either in the sea or on exposed land. In favor of the latter scenario, the emergence of life on land [479] implies that photochemically formed molecules underwent concentration and complexification in ponds, eventually becoming living cells, before being transported by rivers to the sea, which the cells then colonized. Other suggested favorable

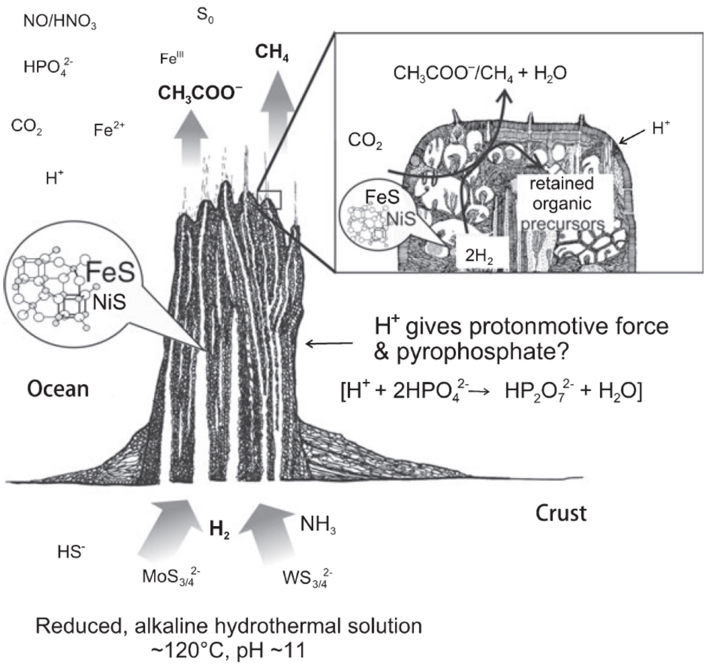

Figure 4.23. The emergence of protolife in a hydrothermal vent [482] John Wiley & Sons (2010).

environments for prebiotic chemistry and the emergence of protocells and of the first living entities include metal-rich, alkaline hydrothermal vents (see figure 4.23) [469, 480–482], or hydrothermal sediments [483].

What is of interest here is the relationship between the proposed locations for the emergence of life and the flux of solar radiation during the Hadean Earth. As was noted above, high-energy radiation is necessary to initiate prebiotic processes to overcome a kinetic barrier that is high enough to prevent reversibility of the chemical reactions. Without free oxygen in the atmosphere and without an ozone layer, radiation could reach the surface of the Earth with few obstacles. Some attenuation could have occurred due to volcanic and impact dust in the atmosphere, but these episodes would have been sporadic and localized, although supervolcano eruptions can eject significant quantities of dust into all levels of the atmosphere, remaining there for many years. We also noted above the possibility of photochemistry resulting in an organic haze if the atmospheric CH_4 concentrations were too high. Obviously, energetic radiation, especially at wavelengths below 103 nm and providing energies greater than 100 kJ $mol^{-1,}$ would have illuminated exposed landmasses, and this is one of the reasons why prebiotic processes may have occurred in such environments on land. However, landmasses on the Hadean Earth must have been restricted to sparse emerged volcanic edifices, such as Iceland or Kerguelen today. Indeed, the prevailing hypothesis is that early continents resembled submerged oceanic plateaus, like the Kerguelen Islands today.

Unfortunately, the lack of preservation of Hadean crust excludes a better understanding of the geological nature of the early Earth as the habitat for the emergence of life. However, if the Palaeoarchaean Earth is, to a certain extent, a representative of the Hadean Earth in this respect, we can make some hypotheses about the environments necessary for the emergence of life, also taking into account the influence of radiation.

As a result of tectonic cycling and impact 'gardening' of the early Earth, its geological record is very patchy and restricted to a few remnants, and even fewer well-preserved remnants), of volcanic and sedimentary rocks that were formed on top of the early protocontinents. There is little evidence for sediment deposition at depths much beyond 100 m or so, while much evidence exists for the deposition in shallow waters, in an offshore to littoral setting. Also, rare deposits of subaerial sedimentation exist in the form of alluvial fans. The question is to what depth the high-energy radiation, necessary for initiating prebiotic chemistry, could penetrate in the early oceans? Radiation penetration in seawater is dependent on wavelength, as well as on the transparency of the waters. During eruption and impacts, large amounts of detritus would decrease transparency, but these events would have been relatively spaced out in time. There would have been periods in between such events that may have been relatively long (perhaps 10^4–10^6 years), when little detritic sedimentation existed in the shallow seas and lagoons, but these were also periods of quieter hydrothermal activity. Nevertheless, a certain fraction of the high-energy radiation required for initial prebiotic processes is strongly attenuated in water. Wavelengths between 200 and 1000 nm (i.e., mid-UV to visible light) are absorbed in the upper 1 cm of the ocean, while higher and lower wavelengths can penetrate to deeper depths. In practice, the longer wavelengths of visible light can penetrate to depths of roughly 50 m in seawater.

Does this mean that locations such as hydrothermal vents and their associated sediments were not conducive to initiating prebiotic chemistry? This is not necessarily the case. Organic molecules could have rained down through the atmosphere to the surface of the Earth, both to ocean and exposed landmasses. At the surface of the ocean, soluble organic matter could have been dissolved and be reprocessed by photochemistry in the upper layers of the water. The products would then sink through the water column to the seafloor and be incorporated into the chemical processing taking place in and around hydrothermal vents. Note that these vents occurred at all depths on the sea floor, including the very shallow near-shore and littoral regions.

While high-energy radiation is apparently required for initiating prebiotic chemistry [467, 479], it is also destructive to organic molecules, especially those that are compositionally and structurally more complex, as was demonstrated by exposure experiments carried out in space and in the laboratory [471, 484]. Thus, at a certain point in the chemical processes leading to the emergence of life, the molecular building bricks of life needed to be protected from radiation. Environments below the penetration depth of UV or in subaqueous sediments or low-temperature hydrothermal vents would be protected. Subaerial environments, on the other hand, would not provide such protection. Since wetting/drying cycles are one of the ways to concentrate prebiotic molecules and promote their incorporation (e.g., RNA inside lipid vesicles), the radiation flux could be a severe hindrance to the survivability of the molecules (see, e.g., [485]). One point of interest regarding the effects of the fainter young Sun is that, while glaciated conditions have been proposed for the Hadean period because of the uncertainties regarding the warming of the early atmospheres, the ancient sedimentary horizons demonstrate absolutely no evidence for glaciated conditions. The formations preserved were largely deposited in shallow waters close to the coastline. These would have been the first areas to be glaciated. On the contrary, there is ample evidence for the presence of liquid water, with waves and storms effortlessly interacting with the underlying sediments [483]. Indeed, volcanic activity accompanied by abundant hydrothermal activity ensured that, at the rock/water interface, temperatures must have been rather warm, up to 70 °C [486].

4.6.2.5 How does the solar emission vary on different timescales?

A specific feature of the Sun's internal dynamo and the associated magnetic activity is the modulation of its output. This modulation manifests itself on many different timescales, with the dominant one being the 11-year Schwabe or sunspot cycle. The 400 year long international sunspot number record represents the longest direct time series of solar activity indices available to us and provides an important insight into the existence of some longer term cyclic variations. Indirect methods such as mass spectrometry now also allow us to infer levels of solar activity throughout the duration of the Holocene using the measurement of ^{14}C and ^{10}Be isotopes inside tree rings and ice cores. These records have enabled the identification of longer but more intermittent cyclic variations. These longer cycles include the Gleissberg cycle (80–150 years), the de Vries/Suess cycle (210 years), and the Hallstatt cycle, a quasi-periodic cycle of 2000–2400 years. In general, these datasets indicate that during the last eleven millennia, the Sun has spent most of its time at moderate activity levels, between 15%–20% in so-called Grand minima and around 10%–15% of its time in Grand maxima [487]. The Grand minima appear to be the result of a stochastic and distinct dynamo mode, persisting for both short and longer periods, while the origin and frequency of Grand maxima remains a topic of considerable debate [487]. On shorter timescales a quasi-biennial (2-year) oscillation in various solar activity proxies has also been identified by many authors [488, 489] as well as in helioseismic frequency measurements, that reveal an oscillation with 11 years of periodicity [490].

A direct consequence of these cycles of magnetic activity is the episodic release of stored magnetic energy and its conversion into other forms such as kinetic energy, accelerated particles, electromagnetic radiation, and heat. This energy release manifests itself observationally as solar flares, during which rapid and localized enhancements across the electromagnetic spectrum occur, accompanied by the acceleration of charged particles. The largest enhancements occur in the UV, EUV, and x-ray regions of the spectrum and since these photons travel at the speed of light their impacts are felt rapidly (8 min later) inside the Earth's magnetosphere, ionosphere, and upper atmosphere. The frequency of flares follows well, though not perfectly, the sunspot cycle, and some of the largest flares of previous solar cycles occurred during the declining phases of activity. The reason for this has been an open question for many years, and is still not well understood.

Often associated with some flares, but sometimes occurring with no obvious flares or other signatures at all, are coronal mass ejections (CMEs). CMEs are 'bubbles' of plasma and magnetic fields sporadically expelled from the Sun's atmosphere. In contrast to the flares that often accompany them, their effects on the solar surface are typically global and their speeds range from a few hundred km s^{-1} to over 2000 km s^{-1} for the fastest events. While this is significantly slower than the speed of light, one of the key properties of CMEs is their speed relative to the ambient solar wind. If their speed is larger, they can drive shock waves that accelerate ambient charged particles to high energies, increase their global extent, and the magnetic field inside the bubbles. In particular, the orientation of the CME magnetic field with respect to the Earth's magnetic field is critical in determining the scale of effects within the magnetosphere, the ionosphere, and the upper atmosphere. When a wide and fast CME with Southward directed internal magnetic field arrives at Earth, several conditions may be satisfied for the onset of a geomagnetic storm. The magnetosphere is pushed inward and the position of the magnetopause (the so-called stand-off distance where inward pressure of solar wind and outward magnetic pressure of the Earth compensate) is reduced. The interaction of oppositely directed magnetic fields leads to a process called reconnection, driving large field-aligned currents that open up the magnetic field at the polar cap, thereby allowing energetic particles from the CME and solar wind to enter the atmosphere and to interact with ambient particle species. The combination of all of these effects on the environment near Earth is described by the term 'space weather'. The strength of the ensuing geomagnetic storm following the impact of a CME depends not only on the properties of the CME, but also on the Earth's magnetic field and on the conditions within the magnetosphere. This significantly complicates making a reliable prediction for the level of likely disruption, requiring detailed understanding of the coupling of many scales between the site of reconnection in the solar atmosphere and the Earth. For our digitally reliant and globally connected societies, a large space weather event represents a genuine risk of significant disruption to both ground- and space-based infrastructures, with associated safety, security, and economic implications. For this reason, space weather features on the national risk registers of many nations. Space weather is discussed in more detail in Chapter 5.

In addition to the solar energetic particles which are accelerated during flares and CMEs, the Earth is, and has always been continuously bombarded by cosmic rays that are produced outside our Solar System with typically much higher energy than the particles produced by the Sun. The flux of cosmic rays that reaches the Earth is, however, modulated by the solar magnetic field with the highest cosmic ray fluxes reaching Earth during periods when the solar magnetic field is weakest. The variation in galactic cosmic ray (GCR) flux depends also on the polarity of the global solar field and exhibits a clear lag with respect to sunspot number, which is also polarity dependent.

Stellar irradiance is a key component in almost every step in the formation and evolution of planetary systems, even at the very earliest stages. For example, Gudel [491] describes how the interaction between the circumstellar gas and dust disk during star formation influences the disk structure and subsequent planetary formation, and how strong temperature gradients induced by x-ray heating of the disk can produce the complex chemical conditions that are required for the formation of planets and planetary atmospheres. The current spectral distribution of solar irradiance is dominated by wavelengths in the optical and infrared regions of the electromagnetic spectrum [492]. The total solar irradiance of the early Sun is estimated to be about 75% of the current value [493], but with a significantly higher UV component. The radiation that reaches the Earth's surface also depends strongly on atmospheric attenuation, highlighting the importance of the coupling between external radiation and atmospheric composition and evolution. Today, atmospheric attenuation provides us considerable protection from harmful UV radiation. All UVC (<280 nm), most UVB (280–320 nm), and some UVA (320–400 nm) is blocked by the existence of a stratospheric ozone. However, this was not the case for early Earth, which had a rather different atmospheric chemistry and no Ozone.

Today it is well established that the total solar irradiance (TSI) varies by only about 0.1% over the period of the sunspot cycle, but evidence accumulated over the last 25 years indicates that that 0.1% total variation masks much larger wavelength-dependent changes as large as 10% in the UV [494]. As we have seen from the previous discussion on the emergence of life, variations in UV radiation can have significant impacts on radiative heating, on the production of ozone, and on other chemical reactions in the middle atmosphere.

Our understanding of stellar evolution, coupled with observational evidence from solar mass stars (or solar analogues), such as T-Tauri stars, suggests that in its pre-main sequence and early main-sequence evolutionary phase our Sun must have rotated much more rapidly than it does today. The consequence of that rapid rotation was a much stronger internal magnetic dynamo which would have led to increased magnetic activity. The manifestations of that higher magnetic activity would likely have included larger sunspots covering more of the solar disk, increased frequency of flares and CMEs, and a stronger solar wind. Additional evidence to support the presence of elevated solar activity comes from mineralogical studies, which can only be explained by much higher fluxes of energetic protons bombarding the Earth in its early evolution.

How can we tell from the study of solar analogues what the past evolution of our Sun looked like? Soderblom *et al* [495] describe the factors that control the rotation rate for solar mass stars a few 100 Myr into their evolution. They find that the effect of loss of angular momentum induced by a stellar wind on its rotation is primarily determined by age. The rotation rate in turn controls the internal magnetic dynamo and hence the level of magnetic activity. This feature of stellar evolution then provides us with a means to infer the activity history of our own Sun through the study of solar analogues spanning a range of ages back to early main sequence. Such a study is the goal of the 'Sun in Time' program [491, 496]. Gudel [491] also notes that while it is difficult to determine precisely how much more active the young Sun must have been, if it has followed a typical main-sequence behavior, a rotation period that is one order of magnitude shorter than the current period of 25–35 days is likely. This implies an x-ray flux that would be 100 times larger than the present level.

While such studies tell us that the Sun must have been more active in the past, how that activity will evolve in the future is far less understood. Even on relatively short timescales of stellar evolution, the recent extended minimum phase of solar cycle 23 sparked a fierce debate about the likelihood of an impending Grand minimum. It is clear that convection and rotation, and especially differential rotation, are critical to the generation of the global dipolar field, and helioseismic measurements increasingly indicate the additional importance of shear stress both for field amplification and surface dynamo processes [497]. What remains a challenge for all current dynamo models is to accurately reproduce cyclic effects, and since the cyclic variation of magnetic activity correlates with increased radiation and particle output, this is a key area of future research.

4.6.2.6 How does solar variability affect the Earth's climate and life?

The Sun has provided the dominant external energy source for our planet since its formation. On the basis of observations of flaring activity in T-Tauri stars, Canuto *et al* [498] estimated that the young Sun could have emitted 10^4 times more UV radiation in the Archean period than it does today. Meert *et al* [499] argue that this strong UVB radiation was able to reach the Earth's surface before the development of the ozone layer, and must have been a key selection driver in evolutionary terms, favoring the development of organisms capable of burrowing vertically and to develop exoskeletons. UV radiation also promotes vitamin D synthesis and affords protection against some viruses and bacteria, equally important positive evolutionary drivers. Nitrogen is present in all complex biologically important molecules but molecular nitrogen requires 'nitrogen fixation' for those complex compounds to form. High temperatures are required for this fixation to occur which could be created through lightning, shock heating, as well as by solar UV radiation, as discussed by reference [500]. In particular, this work explores how the increased level of solar activity during the Sun's young and rapidly rotating phase could have provided the conditions necessary for this fixation to occur through the increased frequency for the occurrence of so-called 'superflare ejections' that produced elevated levels of high-energy radiation and energetic particles.

One key way in which solar variability continues to affect the Earth is through climate forcing. Historically the only solar forcing term included in climate simulations was total solar irradiance (TSI). However, the growing recognition of the importance of spectral solar irradiance (SSI) variability, as well as energetic particle precipitation, has led to the inclusion of SSI in stratospheric models [501–504]. Solar forcing has additional effects beyond atmospheric heating and ozone production, also influencing the lower atmosphere and oceans, as demonstrated by [502] and more recently by [505]. Matthes *et al* [466] note that in addition to volcanic activity, solar variability is a central external driver of climate variability, and that the relative stability of the sunspot cycle may provide a useful tool for improving the reliability of climate predictions on timescales as long as decades. While the effect of energetic particles produced during solar flares, CMEs, and galactic cosmic rays (GCRs) on ionization levels in the ionosphere and subsequent changes in chemical composition is well established and relatively well understood [506], a significant controversy remains about the influence of energetic particles on cloud production. The process of aerosol formation seeded by energetic ions has been demonstrated in the lab, but Dunne *et al* [507] find that the connection between GCRs and cloud production is not very strong.

4.6.3 The grand challenges

From the above, it is clear that our understanding of the early environment of Earth, the environment(s) in which life must have emerged, is far from being well-defined. One of the greatest problems is the lack of a well-preserved crust older than 3.5 Ga. The high degree of metamorphic alteration of older crustal fragments from Greenland and Canada complicate deciphering of the environmental record, although new *in situ* techniques and new discoveries of geochemical proxies for environmental signatures are starting to overcome the problem of metamorphic overprint. New areas of very ancient crust may yet be discovered under the Antarctic ice cover, although the scenario of investigating emerging life at places with molten ice is unpopular from the point of view of global warming. Perhaps other deposits hosting reworked ancient zircon crystals dating back to more than 4 Ga may be brought to light in the future. Geochemical signatures in ancient zircons have been used as proxies to infer the presence of relatively low-temperature hydrothermal fluids and hydrated crust, taken as evidence for the existence of oceans at already 4.3 Ga [508, 509], but it was later recognized that the zircons were actually much younger, and date from the Eo-Palaeoarchaean [510].

One important question is, to what extent can we use the well-preserved Palaeoarchaean rocks as a proxy for the Hadean? Did protocontinents exist in the Hadean, as in the Palaeoarchaean? On the basis of the abundant inherited zircon crystals supposedly inherited from the Hadean era [511], it was originally believed that they would have been common, but the redating of the zircons showed them to be Eo-Palaeoarchaean in age and not Hadean. The presence of a felsic crust, i.e., a fractionated crust formed by water interaction with an ultramafic, basaltic so-called protocontinental crust, is testified by geochemical signatures in younger, Eo-Palaeoarchaean rocks [512]. It may not be possible to realistically determine how abundant the Hadean protocontinents were. Only very rare enclaves of such crust have survived even to the Palaeoarchaean. The presence or not of protocontinents and exposed landmasses is of relevance for early prebiotic chemistry and, to a large extent, to the emergence of life. What was the influence of radiation on subaerial environments in the Hadean and how did it affect prebiotic chemistry and the emergence of life?

The cooling of the Earth after the last, giant Moon-forming impact, the timing of the outgassing of the atmosphere, the condensation of the oceans, the heat flux from the mantle, and the early solar emissivity are all closely interlinked. On the early Earth there was a more or less continuous influx of extraterrestrial organic matter, as well as small organic molecules produced in the chemical reaction of hydrothermal fluids with crustal rocks by the chemical Fischer–Tropsch process. Furthermore, organic molecules were formed in the atmosphere by chemical Strecker synthesis. After the Moon-forming impact, and the subsequent ubiquitous volcanic and hydrothermal activity as the Earth was cooling down, the earliest moment when prebiotic chemistry could have been initiated must have been when oceans condensed and when water temperatures became colder than 120 °C. As we have argued above, the temperature must also have been controlled by the composition of the atmosphere. For example, what was the partial pressure of CO_2 during the Hadean, what were the greenhouse gases? Was there an organic haze because of photo-dissociation of CH_4 in the atmosphere? What was the amount of cloud cover and its effects on albedo? This information can be approached through modeling and comparative planetology and hopefully the use of geochemical proxies for mantle temperature and composition. The same holds for the volatile composition of the early Earth's outer envelope.

There is an active discussion going on concerning the origin of the Earth's volatile layer. Were the planetesimals that formed the Earth rich in volatiles that subsequently outgassed, as suggested by [450], or did a majority of them arrive in a late veneer of extraterrestrial origin [513]? Today it is believed that the Earth's volatile inventory is of mixed origin, but there is much room for improvement in our understanding [514].

These considerations are of enormous relevance to the search for extraterrestrial life. Finding traces of life elsewhere, either on Mars, on icy satellites such as Europa or Enceladus, or on exoplanets using biosignature gaseous combinations, will underline the hypothesis that, under certain environmental conditions, chemistry and physics may natuarally lead to biology, that life is a indeed natural consequence of physics and chemistry, 'vital dust' as was proposed by the Nobel Prize laureate Christian deDuve in 1995. The future NASA Europa Clipper and ESA JUICE missions to Jupiter's moon Europa and planned missions to Saturn's moon Enceladus will go a long way in helping us to understand the processes leading to the emergence of life and the importance of solar radiation for prebiotic chemistry, that is provided they find convincing traces of life.

Today the magnetic field of the Earth protects us from radiation originating from the Sun and beyond, but when did the magnetic field appear? The oldest traces of magnetism on Earth are found in rocks as old as 3.42 Ga from the Barberton Greenstone Belt [515, 516]. Previous interpretations of a Hadean magnetic field based on analysis of zircon crystals [517] have been shown to be erroneous [518]. However, there was already a dynamo on Mars that left its imprint in magnetized rocks dating to 4.2 Ga [519]. Mars did not undergo a magma ocean-forming impact as did the Earth and, thus cooled down more rapidly than the Earth. The initiation of a magnetic dynamo requires that the outer core be in a molten state. The magnetic field of Mars was of short duration and rapidly disappeared, with the core cooling rapidly. This contributed largely to the erosion of the volatile envelope of Mars and its climatic degradation. Could a magnetic field have been initiated on Earth much earlier than 3.5 Ga? This question will have to be addressed by modeling and comparative planetology.

It is clear that variations in solar activity have played, and continue to play, a key role in the evolution of the conditions that support life on Earth. While our ability to accurately predict the strength and duration of the solar cycle has made substantial progress in recent decades, dynamo models still struggle to reproduce the observed asymmetry in sunspot number and duration of each solar cycle. A key element of the regeneration of the global magnetic field during every cycle is the polarity reversal, and our vantage point on the Earth–Sun line has severely restricted our ability to obtain good observations of the polar magnetic field throughout the solar cycle. The extended mission phase of Solar Orbiter, launched in February 2020, will provide important breakthroughs when it begins to leave the ecliptic plane. Given that the solar cycle includes two 11-year sunspot cycles and two polarity reversals, it will be important to ensure the continued availability of such measurements beyond the Solar Orbiter's extended mission.

What also still remains beyond our current reach is the reliable identification of the mechanism that triggers flares and CMEs. This is not just needed to improve space weather forecasting tools. It is the fundamental underpinning science that is needed to understand energy release in magnetized plasmas throughout the Universe as well as in the laboratory, including problems as diverse as plasma confinement in tokamaks, the habitability of exoplanets, and the functioning of blazar jets. Progress in this area requires the ability to reliably measure the magnetic field throughout the atmosphere, particularly inside the corona where the energy is released, and to measure corresponding plasma changes that trace where energy is released and transported. Much like the reliable forecasting of space weather events, one of the challenges here is the complex coupling between different scales, which requires observations at both very high spatial and temporal resolution, and over large fields of view. Simulations, particularly data driven, will also be crucial in this respect.

The uncertainties around the solar forcing of climate is another major challenge, both from the perspective of the uncertainty in the forcing term itself, as well as from that in the uncertainty of the simulated response of the climate. Long-term stable and well-calibrated observations in the UV and shorter wavelengths are needed here in order to improve both model validation and simulations. While there is currently significant disagreement over the importance of GCRs in seeding clouds, the effect of energetic particles on climate is an area that needs further exploration, including the impact of variations in the strength of the solar magnetic field that modulates the GCR flux reaching Earth. But the strength of the solar magnetic field may have broader impact. A study by Lockwood [520] found evidence that extreme winter temperatures in the Northern hemisphere may be the result of a solar influence on the occurrence of jet stream 'blocking' events in the Atlantic. While that is both a regional and seasonal effect, it raises the question of how to quantify the importance of longer-term variations in solar magnetic field strength on climate timescales, both regionally and globally. Finally, the greatest uncertainty in our understanding of solar spectral irradiance is in the UV and at even shorter wavelengths where measurements must be made outside the Earth's atmosphere.

References

[1] Crowe M J 2008 The extraterrestrial life debate *Antiquity to 1915: A Source Book* (Notre Dame, IN: University of Notre Dame Press)

[2] Mifflin M 2009 *The Blue Tattoo: The Life of Olive Oatman* (Lincoln: University of Nebraska Press)

[3] Arcichovsky V M 1912 Auf der Suche nach Chlorophyll auf den Planeten *Ann. Inst. Polytechnique DonCesarevitch Alexis a Novotcherkassk* **1** 195–214
[4] Jeans J 1930 *The Universe Around US* (Cambridge: Cambridge University Press)
[5] Lowell P 1895 Evidence of a twilight arc upon the planet Mars *Nature* **52** 401–5
[6] Lowell P 1906 First photographs of the canals of Mars *Proc. R. Soc.* A **77** 132–5
[7] Sinton W M 1957 Spectroscopic evidence for vegetation on Mars *Astrophys. J.* **126** 231
[8] Sinton W M 1959 Further evidence of vegetation on Mars *Science* **130** 1234–7
[9] Rea D G, O'Leary B T and Sinton W M 1965 Mars: the origin of the 3.58- and 3.69-micron minima in the infrared spectra *Science* **147** 1286–8
[10] McCauley J F, Carr M H, Cutts J A, Hartmann W K, Masursky H, Milton D J, Sharp R P and Wilhelms D E 1972 Preliminary Mariner 9 Report on the geology of Mars (A 4. 3) *Icarus* **17** 289–327
[11] Biemann K *et al* 1976 Search for organic and volatile inorganic compounds in two surface samples from the Chryse Planitia region of Mars *Science* **194** 72–6
[12] Glavin D P, Schubert M, Botta O, Kminek G and Bada J L 2001 Detecting pyrolysis products from bacteria on Mars *Earth Planet. Sci. Lett.* **185** 1–5
[13] dos Santos R, Patel M, Cuadros J and Martins Z 2016 Influence of mineralogy on the preservation of amino acids under simulated Mars conditions *Icarus* **277** 342–53
[14] Fornaro T, Boosman A, Brucato J R, ten Kate I L, Siljeström S, Poggiali G, Steele A and Hazen R M 2018 UV irradiation of biomarkers adsorbed on minerals under Martian-like conditions: hints for life detection on Mars *Icarus* **313** 38–60
[15] Garry J R C, ten Kate I L, Martins Z, Nørnberg P and Ehrenfreund P 2006 Analysis and survival of amino acids in Martian regolith analogs *Meteorit. Planet. Sci.* **41** 391–405
[16] Laurent B, Cousins C R, Pereira M F C and Martins Z 2019 Effects of UV-organic interaction and Martian conditions on the survivability of organics *Icarus* **323** 33–9
[17] Poch O, Noblet A, Stalport F, Correia J J, Grand N, Szopa C and Coll P 2013 Chemical evolution of organic molecules under Mars-like UV radiation conditions simulated in the laboratory with the 'Mars organic molecule irradiation and evolution' (MOMIE) setup *Planet. Space Sci.* **85** 188–97
[18] Poch O, Kaci S, Stalport F, Szopa C and Coll P 2014 Laboratory insights into the chemical and kinetic evolution of several organic molecules under simulated Mars surface UV radiation conditions *Icarus* **242** 50–63
[19] Poch O, Jaber M, Stalport F, Nowak S, Georgelin T, Lambert J-F, Szopa C and Coll P 2015 Effect of nontronite smectite clay on the chemical evolution of several organic molecules under simulated Martian surface ultraviolet radiation conditions *Astrobiology* **15** 221–37
[20] Stoker C R and Bullock M A 1997 Organic degradation under simulated Martian conditions *J. Geophys. Res.* **102** 10881–8
[21] ten Kate I L, Garry J R C, Peeters Z, Quinn R, Foing B and Ehrenfreund P 2005 Amino acid photostability on the Martian surface *Meteorit. Planet. Sci.* **40** 1185–93
[22] ten Kate I L, Garry J R C, Peeters Z, Foing B and Ehrenfreund P 2006 The effects of Martian near surface conditions on the photochemistry of amino acids *Planet. Space Sci.* **54** 296–302
[23] Crandall P B, Góbi S, Gillis-Davis J and Kaiser R I 2017 Can perchlorates be transformed to hydrogen peroxide (H_2O_2) products by cosmic rays on the Martian surface? *J. Geophys. Res.: Planets* **122** 1880–92

[24] Pavlov A A, Vasilyev G, Ostryakov V M, Pavlov A K and Mahaffy P 2012 Degradation of the organic molecules in the shallow subsurface of Mars due to irradiation by cosmic rays *Geophys. Res. Lett.* **39** L13202
[25] Benner S A, Devine K G, Matveeva L N and Powell D H 2000 The missing organic molecules on Mars *Proc. Natl Acad. Sci.* **97** 2425–30
[26] Hecht M H *et al* 2009 Detection of perchlorate and the soluble chemistry of Martian soil at the phoenix lander site *Science* **325** 64–7
[27] Kounaves S P *et al* 2010 Wet chemistry experiments on the 2007 Phoenix Mars Scout Lander mission: data analysis and results *J. Geophys. Res.* **115** E00E10-1–E00E10-16
[28] Gaudi B S, Seager S, Mennesson B, Kiessling A, Warfield K and Cahoy K *et al* 2020 The habitable exoplanet observatory (HabEx) mission concept study final report *ArXiv E-Prints* arXiv:2001.06683 (https://jpl.nasa.gov/habex/pdf/HabEx-Final-Report-Public-Release.pdf)
[29] National Academies of Science 2018 *Exoplanet Science Strategy* (Washington, DC: National Academies Press)
[30] Haldane J B S 1929 The origin of life *Ration. Ann.* **148** 3–10
[31] Haldane J B S 1954 The origin of life *New Biol.* **16** 12–27
[32] Oparin A I 1924 Proiskhodenie Zhizni. Moscoksky Rabotichii, Moscow translated by A S Bernal 1967 *The Origin of Life* In: R Carrington *The Origin of Life* (London: Weidenfeld and Nicolson) pp 199–234
[33] Urey H C 1952 *The Planets, Their Origin and Development* (New Haven: Yale University Press)
[34] Miller S L 1953 A production of amino acids under possible primitive Earth conditions *Science* **117** 528–9
[35] Miller S L and Urey H C 1959 Organic compound synthesis on the primitive Earth *Science* **130** 245–51
[36] Kasting J F, Eggler D H and Raeburn S P 1993 Mantle redox evolution and the oxidation state of the Archean atmosphere *J. Geol.* **101** 245–57
[37] Kasting J F and Catling D 2003 Evolution of a habitable planet *Annu. Rev. Astron. Astrophys.* **41** 429–63
[38] Olson J M 2006 Photosynthesis in the Archean era *Photosynth. Res.* **88** 109–17
[39] Walker J C G 1986 Carbon dioxide on the early Earth *Origins Life* **16** 117–27
[40] Zahnle K, Schaefer L and Fegley B 2010 Earth's earliest atmospheres *Cold Spring Harb. Perspect. Biol.* **2** a003467
[41] Cleaves H J, Chalmers J H, Lazcano A, Miller S L and Bada J L 2008 *Orig. Life Evol. Biosph.* **38** 105–15
[42] Plankensteiner K, Reiner H and Schranz B *et al* 2004 Prebiotic formation of amino acids in a neutral atmosphere. by electric discharge *Angew. Chem. Int. Ed. Engl.* **43** 1886–8
[43] Schlesinger G and Miller S L 1983 Prebiotic synthesis in atmospheres containing CH_4, CO and CO_2. II. Hydrogen cyanide, formaldehyde and ammonia. *J. Mol. Evol.* **19** 383–90
[44] Schlesinger G and Miller S L 1983 Prebiotic synthesis in atmospheres containing CH_4, CO, and CO_2. I. Amino acids *J. Mol. Evol.* **19** 376–82
[45] Stribling R and Miller S L 1987 Energy yields for hydrogen cyanide and formaldehyde syntheses: the HCN and amino acid concentrations in the primitive ocean *Orig. Life Evol. Biosph.* **17** 261–73
[46] Furnes H, Banerjee N R, Muehlenbachs K, Staudigel H and de Wit M 2004 Early life recorded in archean pillow lavas *Science* **304** 578–81

[47] Javaux E J, Marshall C P and Bekker A 2010 Organic-walled microfossils in 3.2-billion-year-old shallow-marine siliciclastic deposits *Nature* **463** 934–8
[48] Mojzsis S J, Arrhenius G, McKeegan K D, Harrison T M, Nutman A P and Friend C R L 1996 Evidence for life on Earth before 3,800 million years ago *Nature* **384** 55–9
[49] Rosing M T 1999 ^{13}C-Depleted carbon microparticles in >3700-Ma sea-floor sedimentary rocks from west Greenland *Science* **283** 674–6
[50] Rosing M T and Frei R 2004 U-rich Archaean sea-floor sediments from Greenland—indications of >3700 Ma oxygenic photosynthesis *Earth Planet. Sci. Lett.* **217** 237–44
[51] Schidlowski M 1988 A 3,800-million-year isotopic record of life from carbon in sedimentary rocks *Nature* **333** 313–8
[52] Sugitani K, Grey K and Nagaoka T *et al* 2009 Taxonomy and biogenicity of Archaean spheroidal microfossils (ca. 3.0 Ga) from the Mount Goldsworthy–Mount Grant area in the northeastern Pilbara Craton, Western Australia *Precambrian Res.* **173** 50–9
[53] Wacey D, McLoughlin N and Whitehouse M J *et al* 2010 Two coexisting sulfur metabolisms in a ca. 3400 Ma sandstone *Geology* **38** 1115–8
[54] Westall F *et al* 2006 *Processes on the Early Earth* ed W U Reimold and R L Gibson (Boulder, CO: Geological Society of America) pp 105–31
[55] Westall F *et al* 2011 Implications of in situ calcification for photosynthesis in a ~3.3 Ga-old microbial biofilm from the Barberton greenstone belt, South Africa *Earth Planet. Sci. Lett.* **310** 468–79
[56] Westall F *et al* 2011 Volcaniclastic habitats for early life on Earth and Mars: a case study from ~3.5 Ga-old rocks from the Pilbara, Australia *Planet. Space Sci.* **59** 1093–106
[57] Westall and Matthews, *Grand Challenges 2050* ch 1.3.5
[58] Cleaves H J and Lazcano A 2009 *Chemical Evolution II: From Origins of Life to Modern Society* ed L Zaikowski, J M Friedrich and S R Seidel (New York: Oxford University Press) pp 17–43
[59] Miller S L and Bada J L 1988 Submarine hot springs and the origin of life *Nature* **334** 609–11
[60] Bains W 2004 Many chemistries could be used to build living systems *Astrobiology* **4** 137–67
[61] Morris R V *et al* 2010 Identification of carbonate-rich outcrops on Mars by the Spirit Rover *Science* **329** 421–4
[62] Poulet F, Bibring J-P, Mustard J F, Gendrin A, Mangold N, Langevin Y, Arvidson R E, Gondet B and Gomez C 2005 Phyllosilicates on Mars and implications for early martian climate *Nature* **438** 623–7
[63] Coradini A, Federico C, Forni O and Magni G 1995 Origin and thermal evolution of icy satellites *Surv. Geophys.* **16** 533–91
[64] Néri A, Guyot F, Reynard B and Sotin C 2020 A carbonaceous chondrite and cometary origin for icy moons of Jupiter and Saturn *Earth Planet. Sci. Lett.* **530** 1–10
[65] Roberts J H and Nimmo F 2008 Near-surface heating on Enceladus and the south polar thermal anomaly *Geophys. Res. Lett.* **35** L09201
[66] Sohl F, Choukroun M, Kargel J, Kimura J, Pappalardo R, Vance S and Zolotov M 2010 Subsurface water oceans on icy satellites: chemical composition and exchange processes *Space Sci. Rev.* **153** 485–510
[67] de Marcellus P, Meinert C, Nuevo M, Filippi J J, Danger G, Deboffle D, Nahon L, d'Hendecourt L L S and Meierhenrich U J 2011 Non-racemic amino acid production by

ultraviolet irradiation of achiral interstellar ice analogs with circularly polarized light *Astrophys. J. Lett.* **727** L27
[68] Meinert C, Filippi J-J, Marcellus P D, D'Hendecourt L L S and Meierhenrich U J 2012 *N*-(2-aminoethyl)glycine and amino acids from interstellar ice analogues *ChemPlusChem* **77** 186–91
[69] Modica P, Meinert C, Marcellus P D, Nahon L, Meierhenrich U J and D'Hendecour L L S 2014 Enantiomeric excesses induced in amino acids by ultraviolet circularly polarized light irradiation of extraterrestrial ice analogs: a possible source of asymmetry for prebiotic chemistry *Astrophys. J.* **788** 79
[70] Modica P, Martins Z, Meinert C, Zanda B and D'Hendecourt L L S 2018 The amino acid distribution in laboratory analogs of extraterrestrial organic matter: a comparison to CM chondrites *Astrophys. J.* **865** 41–51
[71] Muñoz Caro G M, Meierhenrich U J, Schutte W A, Barbier B, Arcones Segovia A, Rosenbauer H, Thiemann W H-P, Brack A and Greenberg J M 2002 Amino acids from ultraviolet irradiation of interstellar ice analogues *Nature* **416** 403–6
[72] Myrgorodska I, Meinert C and Martins Z *et al* 2016 Quantitative enantioseparation of amino acids by comprehensive two-dimensional gas chromatography applied to non-terrestrial samples *J. Chromatogr.* A **1433** 131–6
[73] Brinton K L F, Engrand C, Glavin D P, Bada J L and Maurette M 1998 A search for extraterrestrial amino acids in carbonaceous antarctic micrometeorites *Origins Life Evol. Biosphere* **28** 413–24
[74] Clemett S J, Maechling C R, Zare R N, Swan P D and Walker R M 1993 Identification of complex aromatic molecules in individual interplanetary dust particles *Science* **262** 721–5
[75] Clemett S J, Chillier X D F, Gillette S, Zare R N, Maurette M, Engrand C and Kurat G 1998 Observation of indigenous polycyclic aromatic hydrocarbons in 'Giant' carbonaceous antarctic micrometeorites *Origins Life Evol. Biosphere* **28** 425–48
[76] Dartois E *et al* 2013 Ultra carbonaceous antarctic micrometeorites, probing the Solar System beyond the nitrogen snow-line *Icarus* **224** 243–52
[77] Duprat J *et al* 2010 Extreme deuterium excesses in ultracarbonaceous micrometeorites from Central Antarctic Snow *Science* **328** 742–5
[78] Flynn G J, Keller L P, Feser M, Wirick S and Jacobsen C 2003 The origin of organic matter in the Solar System: evidence from the interplanetary dust particles *Geochim. Cosmochim. Acta* **67** 4791–806
[79] Flynn G J, Keller L P, Jacobsen C and Wirick S 2004 An assessment of the amount and types of organic matter contributed to the Earth by interplanetary dust *Adv. Space Res.* **33** 57–66
[80] Glavin D P, Matrajt G and Bada J L 2004 Re-examination of amino acids in Antarctic micrometeorites *Adv. Space Res.* **33** 106–13
[81] Keller L P, Messenger S, Flynn G J, Clemett S, Wirick S and Jacobsen C 2004 The nature of molecular cloud material in interplanetary dust *Geochim. Cosmochim. Acta* **68** 2577–89
[82] Martins Z 2019 Organic molecules in meteorites and their astrobiological significance *Handbook of Astrobiology* (Boca Raton, FL: CRC Press) pp 177–94
[83] Martins Z and Sephton M A 2009 Extraterrestrial amino acids *Amino Acids, Peptides and Proteins in Organic Chemistry* ed A B Hughes (Weinheim: Wiley-VCH Publishers) pp 3–42

[84] Matrajt G, Pizzarello S and Taylor S *et al* 2004 Concentration and variability of the AIB amino acid in polar micrometeorites: implications for the exogenous delivery of amino acids to the primitive Earth *Meteorit. Planet. Sci.* **39** 1849–58

[85] Matrajt G, Muñoz Caro G M and Dartois E *et al* 2005 FTIR analysis of the organics in IDPs: comparison with the IR spectra of the diffuse interstellar medium *Astron. Astrophys.* **433** 979–95

[86] Martins Z 2020 Detection of organic matter and biosignatures in space missions *Astrobiology: Current, Evolving and Emerging Perspectives* ed A Antunes (Poole: Caister Academic Press) pp 53–74

[87] Martins Z, Chan Q H S, Bonal L, King A and Yabuta H 2020 Organic matter in the Solar System—implications for future on-site and sample return missions *Space Sci. Rev.* **216** 54

[88] Chan M A *et al* 2019 Deciphering biosignatures in planetary contexts *Astrobiology* **19** 1075–102

[89] Des Marais D J, Nuth J A III, Allamandola L J, Boss A P, Farmer J D, Hoehler T M, Jakosky B M, Meadows V S, Pohorille A and Spormann A M 2008 The NASA astrobiology roadmap *Astrobiology* **8** 715–30

[90] Wayne R K, Leonard J A and Cooper A 1999 Full of sound and fury: the recent history of ancient DNA *Annual Rev. Ecol. Syst.* **30** 457–77

[91] Glavin D P, Burton A S, Elsila J E, Aponte J C and Dworkin J P 2020 The search for chiral asymmetry as a potential biosignature in our Solar System *Chem. Rev.* **120** 4660–89

[92] Parnell J *et al* 2007 Searching for life on Mars: selection of molecular targets for ESA's Aurora ExoMars Mission *Astrobiology* **7** 578–604

[93] Olson S L, Reinhard C T and Lyons T W 2016 Limited role for methane in the mid-Proterozoic greenhouse *PNAS* **113** 11447–52

[94] Reinhard C T, Olson S L, Schwieterman E W and Lyons T W 2017 False negatives for remote life detection on ocean-bearing planets: lessons from the early Earth *Astrobiology* **17** 287–97

[95] Schwieterman E W, Kiang N Y, Parenteau M N, Harman C E, DasSarma S and Fisher T M *et al* 2018 Exoplanet biosignatures: a review of remotely detectable signs of life *Astrobiology* **18** 663–708

[96] Arney G, Domagal-Goldman S D and Meadows V S 2018 Organic haze as a biosignature in anoxic Earth-like atmospheres *Astrobiology* **18** 311–29

[97] Domagal-Goldman S D, Meadows V S, Claire M W and Kasting J F 2011 Using biogenic sulfur gases as remotely detectable biosignatures on anoxic planets *Astrobiology* **11** 419–41

[98] Seager S, Schrenk M and Bains W 2012 An astrophysical view of Earth-based metabolic biosignature gases *Astrobiology* **12** 61–82

[99] Des Marais D J, Harwit M O, Jucks K W, Kasting J F, Lin D N C, Lunine J I, Schneider J, Seager S, Traub W A and Woolf N J 2002 Remote sensing of planetary properties and biosignatures on extrasolar terrestrial planets *Astrobiology* **2** 153–81

[100] Seager S, Bains W and Hu R 2013 Biosignature Gases in H_2-dominated atmospheres on rocky exoplanets *Astrophys. J.* **777** 95

[101] Segura A, Kasting J F, Meadows V, Cohen M, Scalo J and Crisp D *et al* 2005 Biosignatures from Earth-like planets around M dwarfs *Astrobiology* **5** 706–25

[102] Sousa-Silva C, Seager S, Ranjan S, Petkowski J J, Zhan Z, Hu R and Bains W 2020 Phosphine as a biosignature gas in exoplanet atmospheres *Astrobiology* **20** 235–68

[103] Zhan Z, Seager S, Petkowski J J, Sousa-Silva C, Ranjan S, Huang J and Bains W 2021 Assessment of isoprene as a possible biosignature gas in exoplanets with anoxic atmospheres arXiv:2103.14228 [astro-ph.EP] *Astrobiology*
[104] Huang J, Seager S, Petkowski J, Ranjan S and Zhan Z 2021 Assessment of ammonia as a biosignature gas in exoplanet atmospheres *Astrobiology* **22** 171–91
[105] Seager S, Bains W and Hu R 2013 A biomass-based model to estimate the plausibility of exoplanet biosignature gases *Astrophys. J.* **775** 104
[106] Grenfell J L 2018 Atmospheric biosignatures *Handbook of Exoplanets* ed H J Deeg and J A Belmonte (Cham: Springer International Publishing) pp 1–14
[107] Kaltenegger L 2017 How to characterize habitable worlds and signs of life *Annu. Rev. Astron. Astrophys.* **55** 433–85
[108] Morowitz H and Sagan C 1967 Life in the clouds of Venus? *Nature* **215** 1259
[109] Seager S, Petkowski J J, Gao P, Bains W, Bryan N C, Ranjan S and Greaves J 2021 The venusian lower atmosphere haze as a depot for desiccated microbial life: a proposed life cycle for persistence of the venusian aerial biosphere *Astrobiology* **21** 1206–23
[110] Cavalazzi B *et al* 2019 The dallol geothermal area, northern Afar (Ethiopia)—an exceptional planetary field analog on earth *Astrobiology* **19** 553–78
[111] Kotopoulou E, Delgado Huertas A, Garcia-Ruiz J M, Dominguez-Vera J M, Lopez-Garcia J M, Guerra-Tschuschke I and Rull F 2018 A polyextreme hydrothermal system controlled by iron: the case of Dallol at the Afar Triangle *ACS Earth Sp. Chem.* **3** 90–9
[112] Limaye S S, Mogul R, Smith D J, Ansari A H, Słowik G P and Vaishampayan P 2018 Venus' spectral signatures and the potential for life in the clouds *Astrobiology* **18** 1181–98
[113] Oyama V I *et al* 1980 Pioneer Venus gas chromatography of the lower atmosphere of Venus *J. Geophys. Res. Sp. Phys.* **85** 7891–902
[114] Mukhin L M *et al* 1982 VENERA-13 and VENERA-14 gas chromatography analysis of the Venus atmosphere composition *Sov. Astron. Lett.* **8** 216–8
[115] Rimmer P B, Jordan S, Constantinou T, Woitke P, Shorttle O, Paschodimas A and Hobbs R 2021 Hydroxide salts in the clouds of Venus: their effect on the sulfur cycle and cloud droplet pH *Planet. Sci. J.* arXiv:2101.08582
[116] Greaves J S, Bains W, Petkowski J J, Seager S, Sousa-Silva C and Ranjan S *et al* 2021 On the robustness of phosphine signatures in Venus' clouds *arXiv preprint* arXiv:2012.05844
[117] Greaves J S, Richards A, Bains W, Rimmer P B, Clements D L and Seager S *et al* 2021 Recovery of spectra of phosphine in Venus' clouds *arXiv preprint* arXiv:2104.09285
[118] Greaves J S *et al* 2020 Phosphine gas in the cloud decks of Venus *Nat. Astron.* **5** 655–64
[119] Mogul R, Limaye S S, Way M J and Cordova J A 2021 Venus' mass spectra show signs of disequilibria in the middle clouds *Geophys. Res. Lett.* **48** e2020GL091327
[120] Surkov Y A, Andrejchikov B M and Kalinkina O M 1973 On the content of ammonia in the Venus atmosphere based on data obtained from Venera 8 automatic station *Akad. Nauk SSSR Dokl.* **213** 296–8
[121] Colin L 1980 The pioneer Venus program *J. Geophys. Res.* **85** 7575
[122] Sagdeev R Z *et al* 1986 The VEGA Venus balloon experiment *Science* **231** 1407
[123] Archer P D *et al* 2014 Abundances and implications of volatile-bearing species from evolved gas analysis of the Rocknest aeolian deposit, Gale Crater, Mars *J. Geophys. Res.: Planets* **119** 237–54
[124] Freissinet C *et al* MSL Science Team 2015 Organic molecules in the Sheepbed Mudstone, Gale Crater, Mars *J. Geophys. Res.: Planets* **120** 495–514

[125] Glavin D P *et al* 2013 Evidence for perchlorates and the origin of chlorinated hydrocarbons detected by SAM at the Rocknest aeolian deposit in Gale Crater *J. Geophys. Res.: Planets* **118** 1955–73

[126] Leshin L A *et al* 2013 Volatile, isotope, and organic analysis of Martian fines with the Mars curiosity rover *Science* **341** 1238937-1–1239

[127] Ming D W *et al* 2014 Volatile and organic compositions of sedimentary rocks in Yellowknife Bay, Gale Crater, Mars *Science* **343** 1245267-1–15

[128] Vago J L *et al* 2017 Habitability on early Mars and the search for biosignatures with the ExoMars Rover *Astrobiology* **17** 471–510

[129] Liuzzi G *et al* NOMAD Team 2019 Methane on Mars: new insights into the sensitivity of CH_4 with the NOMAD/ExoMars spectrometer through its first in-flight calibration *Icarus* **321** 671–90

[130] Mumma M J, Villanueva G L, Novak R E, Hewagama T, Bonev B P, DiSanti M A, Mandell A M and Smith M D 2009 Strong release of methane on Mars in northern summer 2003 *Science* **323** 1041–5

[131] Formisano V, Atreya S, Encrenaz T, Ignatiev N and Giuranna M 2004 Detection of methane in the atmosphere of Mars *Science* **306** 1758–61

[132] Geminale A, Formisano V and Sindoni G 2011 Mapping methane in Martian atmosphere with PFS-MEX data *Planet. Space Sci.* **59** 137–48

[133] Fonti S and Marzo G A 2010 Mapping the methane on Mars *Astron. Astrophys.* **512** A51–7

[134] Fonti S, Mancarella F, Liuzzi G, Roush T L, Chizek Frouard M, Murphy J and Blanco A 2015 Revisiting the identification of methane on Mars using TES data *Astron. Astrophys.* **581** A136-1–A136-11

[135] Webster C R *et al* 2015 Mars methane detection and variability at Gale crater *Science* **347** 415–7

[136] Webster C R *et al* 2018 Background levels of methane in Mars' atmosphere show strong seasonal variations *Science* **360** 1093–6

[137] Korablev O, Vandaele A C and Montmessin F *et al* 2019 No detection of methane on Mars from early ExoMars Trace Gas Orbiter observations *Nature* **568** 517–20

[138] Iess L *et al* 2014 The gravity field and interior structure of Enceladus *Science* **344** 78–80

[139] Thomas P C, Tajeddine R, Tiscareno M S, Burns J A, Joseph J, Loredo T J, Helfenstein P and Porco C 2016 Enceladus's measured physical libration requires a global subsurface ocean *Icarus* **264** 37–47

[140] Postberg F *et al* 2018 Macromolecular organic compounds from the depths of Enceladus *Nature* **558** 564–8

[141] Waite J H *et al* 2006 Cassini ion and neutral mass spectrometer: Enceladus plume composition and structure *Science* **311** 1419–22

[142] Grasset O *et al* 2013 JUpiter ICy moons Explorer (JUICE): an ESA mission to orbit Ganymede and to characterise the Jupiter system *Planet. Space Sci.* **78** 1–21

[143] Kissel J and Krueger F R 1987 The organic component in dust from comet Halley as measured by the PUMA mass spectrometer on board Vega 1 *Nature* **326** 755–60

[144] Cody G D *et al* 2008 Quantitative organic and light-element analysis of comet 81P/Wild 2 particles using C-, N-, and O-μ-XANES *Meteorit. Planet. Sci.* **43** 353–65

[145] Sandford S A *et al* 2006 Organics captured from Comet 81P/Wild 2 by the Stardust Spacecraft *Science* **314** 1720–4

[146] De Gregorio B T, Stroud R M, Cody G D, Nittler L R, David Kilcoyne A L and Wirick S 2011 Correlated microanalysis of cometary organic grains returned by Stardust *Meteor. Planet. Sc.* **46** 1376–96
[147] Elsila J E, Glavin D P and Dworkin J P 2009 Cometary glycine detected in samples returned by Stardust *Meteorit. Planet. Sci.* **44** 1323–30
[148] Fray N *et al* 2016 High-molecular-weight organic matter in the particles of comet 67P/ Churyumov–Gerasimenko *Nature* **538** 72–4
[149] Capaccioni F 2015 The organic-rich surface of comet 67P/Churyumov–Gerasimenko as seen by VIRTIS/Rosetta *Science* **347** aaa0628-1–aaa4
[150] Goesmann F *et al* 2015 Organic compounds on comet 67P/Churyumov–Gerasimenko revealed by COSAC mass spectrometry *Science* **349** aab0689-1–aab4
[151] Wright I P, Sheridan S, Barber S J, Morgan G H, Andrews D J and Morse A D 2015 CHO-bearing organic compounds at the surface of 67P/Churyumov–Gerasimenko revealed by Ptolemy *Science* **349** aab0673-1–aab3
[152] Altwegg K *et al* 2016 Prebiotic chemicals—amino acid and phosphorus—in the coma of comet 67P/Churyumov–Gerasimenko *Sci. Adv.* **2** e1600285–e15
[153] Nakamura T *et al* 2011 Itokawa dust particles: a direct link between S-type asteroids and ordinary chondrites *Science* **333** 1113–6
[154] Uesugi M *et al* 2014 Sequential analysis of carbonaceous materials in Hayabusa-returned samples for the determination of their origin *Earth, Planets Space* **66** 102-1–102-11
[155] Wada K *et al* 2018 Asteroid Ryugu before the Hayabusa2 encounter *Prog. Earth Planet. Sci.* **5** 82 1–30
[156] Naraoka H 2023 Soluble organic molecules in samples of the carbonaceous asteroid (162173) Ryugu *Science* **379** abn9033
[157] Yabuta H 2023 Macromolecular organic matter in samples of the asteroid (162173) Ryugu *Science* **379** eabn9057
[158] Lauretta D S *et al* 2017 OSIRIS-REx: sample return from asteroid (101955) Bennu *Space Sci. Rev.* **212** 925–84
[159] Gardner J P, Mather J C, Clampin M, Doyon R, Greenhouse M A, Hammel H B, Hutchings J B, Jakobsen P, Lilly S J and Long K S *et al* 2006 The James Webb space telescope *Space Sci. Rev.* **123** 485–606
[160] Seager S and Sasselov D D 2000 Theoretical transmission spectra during extrasolar giant planet transits *Astrophys. J.* **537** 916
[161] Gillon M, Triaud A H M J, Demory B-O, Jehin E, Agol E and Deck K M *et al* 2017 Seven temperate terrestrial planets around the nearby ultracool dwarf star TRAPPIST-1 *Nature* **542** 456–60
[162] Turbet M, Bolmont E, Bourrier V, Demory B O, Leconte J, Owen J and Wolf E T 2020 A review of possible planetary atmospheres in the TRAPPIST-1 system *Space Sci. Rev.* **216** 100
[163] Gilmozzi R and Spyromilio J 2007 The European extremely large telescope (E-ELT) *The Messenger* **127** 11 https://ui.adsabs.harvard.edu/abs/2007Msngr.127…11G
[164] Sanders G H 2013 The Thirty Meter Telescope (TMT): an international observatory *J. Astrophys. Astron.* **34** 81–6
[165] Bernstein R A, McCarthy P J, Raybould K, Bigelow B C, Bouchez A H and Filgueira J M *et al* 2014 Overview and status of the Giant Magellan Telescope project *Ground-based and Airborne Telescopes* V **vol 9145** (Bellingham, WA: SPIE) p 91451C

[166] Johns M, McCarthy P, Raybould K, Bouchez A, Farahani A, Filgueira J, Jacoby G, Shectman S and Sheehan M 2012 Giant Magellan telescope: overview *Ground-based and Airborne Telescopes IV* (Bellingham, WA: SPIE) vol 8444 p 84441H

[167] Anglada-Escudé G, Amado P J, Barnes J, Berdiñas Z M, Butler R P and Coleman G A L *et al* 2016 A terrestrial planet candidate in a temperate orbit around Proxima Centauri *Nature* **536** 437–40

[168] Spitzer L 1962 The beginnings and future of space astronomy *Am. Sci.* **50** 473–84

[169] Seager S, Turnbull M, Sparks W, Thomson M, Shaklan S B, Roberge A, Kuchner M, Kasdin N J, Domagal-Goldman S and Cash W *et al* 2015 The Exo-S probe class starshade mission *Tech. Instrum. Detect. Exopl. VII* **9605** 96050W

[170] LUVOIR Team 2019 The LUVOIR Mission Concept Study Final Report *ArXiv E-Prints* arXiv:191206219T https://asd.gsfc.nasa.gov/luvoir/reports/LUVOIR_FinalReport_2019-08-26.pdf

[171] Feng F, Tuomi M, Jones H R A, Barnes J, Anglada-Escudé G, Vogt S S and Butler R P 2017 Color difference makes a difference: four planet candidates around τ ceti *Astron.* J **154** 135

[172] Tuomi M, Jones H R A, Jenkins J S, Tinney C G, Butler R P and Vogt S S *et al* 2013 Signals embedded in the radial velocity noise. Periodic variations in the τ Ceti velocities *Astron. Astrophys.* **551** A79

[173] Chan Q H S, Stroud R, Martins Z and Yabuta H 2020 Concerns of organic contamination for sample return space missions *Space Sci. Rev.* **216** 56

[174] Dhingra R D *et al* 2019 Observational evidence for summer rainfall at Titan's north pole *Geophys. Res. Lett.* **46** 1205–12

[175] Poch O, Coll P, Buch A, Ramírez S I and Raulin F 2012 Production yields of organics of astrobiological interest from H_2O–NH_3 hydrolysis of Titan's tholins *Planet. Space Sci.* **61** 114–23

[176] Eshleman V R 1979 Gravitational lens of the Sun: its potential for observations and communications over interstellar distances *Science* **205** 1133–5

[177] Landis G A 2017 Mission to the gravitational focus of the Sun: a critical analysis *55th AIAA Aerospace Sciences Meeting* (American Institute of Aeronautics and Astronautics)

[178] Turyshev S G, Shao M, Toth V T, Friedman L D, Alkalai L and Mawet D *et al* 2020 Direct multipixel imaging and spectroscopy of an exoplanet with a solar gravity lens mission *ArXiv E-Prints* 2020 arXiv:200211871T

[179] Licklider J C R 1960 Man-computer symbiosis *IRE Trans. Hum. Factors Electron.* **HFE-1** 4–11

[180] Schwab K 2021 The Fourth Industrial Revolution: What It Means, How To Respond *Word Economic Forum* (https://weforum.org/agenda/2016/01/the-fourth-industrialrevolution-what-it-means-and-how-to-respond/)

[181] Wiener N 1948 *Cybernetics: or the Control and Communication in the Animal and the Machine* (New York: Wiley)

[182] Ferrucci D *et al* 2010 Building Watson: an overview of the Deepqa project *AI Mag.* **31** 59–79

[183] Boden M 2014 GOFAI *The Cambridge Handbook of Artificial Intelligence* ed F Keith and W M Ramsey (Cambridge: Cambridge University Press)

[184] Newell A and Simon H A 1976 Computer science as empirical inquiry: symbols and search (ACM 1975 Turing award lecture) *Commun. ACM* **19** 113–26

[185] Franklin S 2014 History, motivations, and core themes *The Cambridge Handbook of Artificial Intelligence* ed F Keith and W M Ramsey (Cambridge: Cambridge University Press)
[186] Feldman R A and Ballard D H 1982 Connectionist models and their properties *Cogn. Sci.* **6** 205–54
[187] James W 1890 *The Principles of Psychology* **vol 1** (Cambridge, MA: Harvard University Press)
[188] Little W A 1974 The existence of persistent states in the brain *Math. Biosci.* **19** 101–20
[189] Hopfield J J 1982 Neural neural network and physical systems with emergent collective computational abilities *Proc. Natl Acad. Sci.* **79** 2554–88
[190] Hinton G E and Sejnowski T J 1986 Learning and relearning in Boltzmann machines *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* ed D E Rumelhart, J L McClelland and The PDP Research Group (Cambridge, MA: MIT Press) pp 282–317
[191] Abbas A, Sutter D, Zoufal C, Lucchi A, Figalli A and Woerner S 2021 The power of quantum neural networks *Nat. Comput. Sci.* **1** 403–9
[192] Dunjko V and Briegel H J 2018 Machine learning and artificial intelligence in the quantum domain: a review of recent progress *Rep. Prog. Phys.* **81** 074001
[193] McCulloch W S and Pitts W 1943 A logical calculus of ideas immanent in nervous activity *Bull. Math. Biophys.* **5** 115–33
[194] Rosenblatt F 1958 The perceptron: a probabilistic model for information storage and organization in the brain *Psychol. Rev.* **65** 386–408
[195] Widrow B and Hoff M E 1960 Adaptive switching circuits *IRE WESCON Convention Record (New York)* pp 96–104
[196] Minsky M and Papert S 1969 *Perceptrons: An Introduction to Computational Geometry* (Cambridge, MA: MIT Press)
[197] Rumelhart D E, Hinton G E and Williams R J 1986 Learning representations by back-propagating errors *Nature* **323** 533–6
[198] Werbos P 1974 Beyond regression: new tools for prediction and analysis in the behavioural sciences *Masters Thesis* Harvard University
[199] Medler D A 1998 A brief history of connectionism *Neural Comput. Surv.* **1** 61–101
[200] Hochreiter S and Schmidhuber J 1997 Long short-term memory *Neural Comput.* **9** 1735–80
[201] LeCun Y, Bottou L, Bengio Y and Haffner P 1998 Gradient-based learning applied to document recognition *Proc. IEEE* **86** 2278–324
[202] Krizhevsky A, Sutskever I and Hinton G 2012 Imagenet classification with deep convolutional neural networks *Advances in Neural Information Processing Systems* ed F Pereira, C J C Burges, L Bottou and K Q Weinberger (Cambridge, MA: MIT Press) vol 25
[203] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (Cambridge, MA: MIT Press)
[204] Schmidhuber J 2014 Deep learning in neural networks: an overview *arXiv preprint* (arXiv:1404.7828 v2)
[205] Mao J, Xu W, Huang Y Y and Yuille A 2015 Deep captioning with multimodal recurrent neural networks (m-RNN) *Proc. of the 3rd Int. Conf. on Learning Representations (ICLR)*
[206] Brown T B *et al* 2020 Language models are few-shot learners arXiv:2005.14165
[207] Danks D 2014 Learning *The Cambridge Handbook of Artificial Intelligence* ed F Keith and W M Ramsey (Cambridge: Cambridge University Press)

[208] Silver D, Huang A, Maddison C J, Guez A, Sifre L and Van Den Driessche G *et al* 2016 Mastering the game of Go with deep neural networks and tree search *Nature* **529** 484–9
[209] Hu Z, Tang J, Wang Z, Zhang K, Zhang L and Sun Q 2018 Deep learning for image-based cancer detection and diagnosis—a survey *Pattern Recognit.* **83** 134–49
[210] Holzinger A 2016 Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Inform.* **3** 119–31
[211] Miotto R, Wang F, Wang S, Jiang X and Dudley J T 2018 Deep learning for healthcare: review, opportunities and challenges *Brief. Bioinform* **19** 1236–46
[212] Holzinger A, Langs G, Denk H, Zatloukal K and Müller H 2019 Causability and explainability of artificial intelligence in medicine *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.* **9** e1312
[213] IEEE Robots 2021 (https://robots.ieee.org/)
[214] Cao H *et al* 2019 Robot-enhanced therapy: development and validation of a supervised autonomous robotic system for autism spectrum disorders therapy *IEEE Rob. Autom Mag.* **26** 49–58
[215] Zhang S, Yao L, Sun A and Tay Y 2019 Deep learning based recommender system: a survey and new perspectives *ACM Comput. Surv. (CSUR)* **52** 1–38
[216] Buolamwini J and Gebru T 2018 Gender shades: intersectional accuracy disparities in commercial gender classification *Conf. on Fairness, Accountability and Transparency Proc. of Machine Learning Research* pp 77–91
[217] Barlow A and Sriskandarajah S 2019 *Artificial Intelligence—Application to the Sports Industry* (https://pwc.com.au/industry/sports/artificial-intelligence-application-to-thesports-industry.pdf)
[218] Hawk-Eye 2021 (https://hawkeyeinnovations.com/)
[219] Shaw D 2017 *How Wimbledon is Using IBM Watson Ai to Power Highlights, Analytics and Enriched Fan Experiences* (https://ibm.com/blogs/watson/2017/07/ibm-watsons-ai-ispowering-wimbledon-highlights-analytics-and-a-fan-experiences/)
[220] AlphaGo 2021 (https://deepmind.com/research/case-studies/alphago-the-story-so-far)
[221] Bartneck C, Belpaeme T, Eyssel F, Kanda T, Keijsers M and Šabanović S 2020 *Human–Robot Interaction: An Introduction* (Cambridge: Cambridge University Press)
[222] Vinanzi S, Patacchiola M, Chella A and Cangelosi A 2019 Would a robot trust you? Developmental robotics model of trust and theory of mind *Phil. Trans. R. Soc. B* **374** 20180032
[223] Sandini G, Sciutti A and Vernon D 2021 Cognitive robotics *Encyclopedia of Robotics* ed M Ang, O Khatib and B Siciliano (Berlin: Springer)
[224] Leite I, Martinho C and Paiva A 2013 Social robots for long-term interaction: a survey *Int. J. Soc. Robot.* **5** 291–308
[225] Mou W, Ruocco M, Zanatto D and Cangelosi A 2020 When would you trust a robot? A study on trust and theory of mind in human–robot interactions *Proc. of RO-MAN2020, 29th IEEE Int. Conf. on Robot and Human Interactive Communication (Naples, August 2020)*
[226] Vinanzi S, Cangelosi A and Goerick C 2021 The collaborative mind: intention reading and trust in human–robot interaction *Science* **24** 102130
[227] Balog M, Gaunt A L, Brockschmidt M, Nowozin S and Tarlow D 2016 DeepCoder: learning to write programs *arXiv preprint* arXiv:1611.01989

[228] Hewitt L B, Le T A and Tenenbaum J B 2020 Learning to infer program sketches Proc. of the 36th Conf. on Uncertainty in Artificial Intelligence
[229] Heaven W D 2020 *OpenAI's New Language Generator GPT-3 Is Shockingly Good—and Completely Mindless* (MIT Technology Review)
[230] Cangelosi A and Schlesinger M 2015 *Developmental Robotics: From Babies to Robots* (Cambridge, MA: MIT Press)
[231] Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M and Bolton A *et al* 2017 Mastering the game of go without human knowledge *Nature* **550** 354
[232] OECDAI 2021 National AI policies & strategies https://oecd.ai/dashboards
[233] Alupo C D, Omeiza D and Vernon D 2022 Realizing the potential of AI in Africa ed M I Aldinhas Ferreira and O Tokhi *Towards Trustworthy Artificial Intelligence Systems, Intelligent Systems, Control and Automation: Science and Engineering* (Berlin: Springer)
[234] Novitske L 2018 *The AI Invasion Is Coming to Africa (and It's a Good Thing)* (Stanford Social Innovation Review)
[235] Pillay N and Access Partnership 2018 *Artificial Intelligence for Africa: An Opportunity for Growth, Development, and Democratization* (White Paper, University of Pretoria)
[236] EUAI 2021 *European Commission Ethics Guidelines for Trustworthy AI* (https://ec.europa.eu/futurium/en/ai-alliance-consultation)
[237] Besaw C and Filitz J 2019 AI and global governance: AI in Africa is a double-edged sword *Tech. rep.* (United Nations University) (https://cpr.unu.edu/ai-in-africa-is-a-doubleedged-sword.html)
[238] Wilson B, Hoffman J and Morgenstern J 2019 Predictive inequity in object detection *arXiv preprint* arXiv:1902.11097
[239] Larson J, Mattu S, Kirchner L and Angwin J 2016 How we analyzed the COMPAS recidivism algorithm *ProPublica* May 2016 9
[240] Wolpert D 2011 *The Real Reason for Brains* (https://youtube.com/watch?v=7s0CpRfyYp8)
[241] Maturana H and Varela F 1987 *The Tree of Knowledge—The Biological Roots of Human Understanding* (Boston, MA and London: New Science Library)
[242] Friston K J 2010 The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11** 127–38
[243] Verschure P F M J 2016 Synthetic consciousness: the distributed adaptive control perspective *Phil. Trans. R. Soc.* B **371** 20150448
[244] Laird J E 2012 *The Soar Cognitive Architecture* (Cambridge, MA: MIT Press)
[245] Anderson J R, Bothell D, Byrne M D, Douglass S, Lebiere C and Qin Y 2004 An integrated theory of the mind *Psychol. Rev.* **111** 1036–60
[246] Sun R 2016 *Anatomy of the Mind: Exploring Psychological Mechanisms and Processes with the Clarion Cognitive Architecture* (Oxford: Oxford University Press)
[247] Langton C G 1997 *Artificial Life: An Overview* (Cambridge, MA: MIT Press)
[248] Adami C 1998 *Introduction to Artificial Life* (New York: Springer)
[249] Aguilar W, Bonfil G S, Froese T and Gershenson C 2014 The past, present, and future of artificial life *Front. Robotics AI* **1** 8
[250] Langton C G 1987 *Artificial Life: The Proc. of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems (Los Alamos, NM, Sept. 1987)* (Addison-Wesley)

[251] Bedau M, McCaskill J, Packard P, Rasmussen S, Green D, Ikegami T, Kaneko K and Ray T 2000 Open problems in artificial life *Artif. Life* **6** 363–76
[252] von Neumann J 1966 *The Theory of Self-Reproducing Automata* ed A W Burks (Champaign, IL: University of Illinois Press)
[253] Shelling T C 1971 Dynamic models of segregation *J. Math. Sociol.* **1** 143–86
[254] Berlekamp E R, Conway J H and Guy R K 1982 *Winning Ways for Your Mathematical Plays Games in Particular* **vol 2** (London: Academic)
[255] Wolfram S 1983 Statistical mechanics of cellular automata *Rev. Mod. Phys.* **55** 601–44
[256] Wuensche A and Lesser M J 1992 *The Global Dynamics of Cellular Automata; An Atlas of Basin of Attraction Fields of One-Dimensional Cellular Automata* (Reading, MA: Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley)
[257] Mitchell M, Hraber P T and Crutchfield J P 1993 Revisiting the edge of chaos: evolving cellular automata to perform computations *Complex Syst.* **7** 89 130
[258] Epstein J M and Axtell R L 1996 *Growing Artificial Societies: Social Science from the Bottom Up* (Cambridge, MA: Brookings Institution Press MIT Press)
[259] Kauffman S A 1969 Metabolic stability and epigenesis in randomly constructed genetic nets *J. Theor. Biol.* **22** 437 467
[260] Aldana-González M, Coppersmith S and Kadanoff L P 2003 Boolean dynamics with random couplings In E Kaplan, J E Marsden and K R Sreenivasan *Perspectives and Problems in Nonlinear Science. A Celebratory Volume in Honor of Lawrence Sirovich, Applied Mathematical Sciences Series* (Berlin: Springer)
[261] Gershenson C 2004 Introduction to random Boolean networks *Workshop and Tutorial Proc., 9th Int. Conf. on the Simulation and Synthesis of Living Systems (ALife IX) (Boston, MA)* ; M Bedau, P Husbands, T Hutton, S Kumar and H Suzuki pp 160–73
[262] Gershenson C 2012 Guiding the self-organization of random Boolean networks *Theory Biosci.* **131** 181–91
[263] Reynolds C W 1987 Flocks, herds, and schools: a distributed behavioral model *Comput. Graph.* **21** 25–34
[264] Sipper M 1998 Fifty years of research on self-replication: an overview *Artif. Life* **4** 237–57
[265] Ray T S 1994 An evolutionary approach to synthetic biology: zen and the art of creating life *Artif. Life* **1** 195–226
[266] Chris Adami and Titus Brown C 1994 Evolutionary learning in the 2D artificial life system 'Avida' *Proc. Artificial Life IV* ed R Brooks and P Maes (Cambridge, MA: MIT Press) pp 377–81
[267] Lenski R E, Ofria C, Collier T C and Adami C 1999 Genome complexity, robustness and genetic interactions in digital organisms *Nature* **400** 661–4
[268] Lenski R E, Ofria C, Pennock R T and Adami C 2003 The evolutionary origin of complex features *Nature* **423** 139–44
[269] Adami C, Ofria C and Collier T C 2000 Evolution of biological complexity *Proc. Natl Acad. Sci.* **97** 4463–8
[270] Fontana W 1991 Algorithmic chemistry: a model for functional self-organization *Artificial Life II* ed C G Langton, C Taylor, J D Farmer and S Rasmussen (Boston, MA: Addison-Wesley) pp 159–202
[271] Dittrich P, Ziegler J and Banzhaf W 2001 Artificial chemistries—a review *Artif. Life* **7** 225–75 2014/07/08
[272] Hiroki S 2008 Swarm chemistry *Artif. Life* **15** 105–14 2014/07/08

[273] Karl S 1994 Evolving 3D morphology and behavior by competition *Artif. Life* **1** 353–72
[274] Jakobi N 1997 Evolutionary robotics and the radical envelope of noise hypothesis *Adapt. Behav.* **6** 325–68
[275] Lipson H and Pollack J B 2000 Automatic design and manufacture of robotic lifeforms *Nature* **406** 974–8
[276] Bongard J, Zykov V and Lipson H 2006 Resilient machines through continuous self-modeling *Science* **314** 1118–21
[277] Brooks R A 1991 Intelligence without representation *Artif. Intell.* **47** 139 –159
[278] Steels L and Brooks R A 1995 *The Artificial Life Route to Artificial Intelligence: Building Embodied, Situated Agents* (New York City: Lawrence Erlbaum Associates)
[279] Beer R D 2014 Dynamical systems and embedded cognition *The Cambridge Handbook of Artificial Intelligence* ed K Frankish and W Ramsey (Cambridge: University Press)
[280] Maes P 1994 Modeling adaptive autonomous agents *Artif. Life* **1** 135–62
[281] Beer R D 1990 *Intelligence as Adaptive Behavior: An Experiment in Computational Neuroethology* (San Diego, CA: Academic)
[282] Braitenberg V 1986 *Vehicles: Experiments in Synthetic Psychology* (Cambridge, MA: MIT Press)
[283] Gershenson C 2004 Cognitive paradigms: which one is the best? *Cognit. Syst. Research* **5** 135–56
[284] Cliff D, Husbands P and Harvey I 1993 Explorations in evolutionary robotics *Adapt. Behav.* **2** 73–110
[285] Nolfi S and Floreano D 2000 *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines* (Cambridge, MA: MIT Press)
[286] Harvey I, Di Paolo E A, Wood R, Quinn M and Tuci E A 2005 Evolutionary robotics: A new scientific tool for studying cognition *Artif. Life* **11** 79–98
[287] Dorigo M *et al* 2004 Evolving self-organizing behaviors for a swarm-bot *Auton. Robots* **17** 223–45
[288] Zykov V, Mytilinaios E, Adams B and Lipson H 2005 Robotics: self-reproducing machines *Nature* **435** 163 164
[289] Halloy J *et al* 2007 Social integration of robots into groups of cockroaches to control self-organized choices *Science* **318** 1155–8
[290] Rubenstein M, Cornejo A and Nagpal R 2014 Programmable self-assembly in a thousand-robot swarm *Science* **345** 795–9
[291] Vásárhelyi G, Virágh C, Somorjai G, Nepusz T, Eiben A E and Vicsek. T 2018 Optimized flocking of autonomous drones in confined environments *Sci. Robot.* **3**
[292] Luisi P L 2002 Toward the engineering of minimal living cells *Anat. Rec.* **268** 208–14
[293] Hanczyc M M, Fujikawa S M and Szostak J W 2003 Experimental models of primitive cellular compartments: encapsulation, growth, and division *Science* **302** 618–22
[294] Rasmussen S, Bedau M A, Chen L, Deamer D, Krakauer D C, Packard N H and Stadler P F 2008 *Protocells: Bridging Nonliving and Living Matter* (Cambridge, MA: MIT Press)
[295] Rasmussen S, Chen L, Deamer D, Krakauer D C, Packard N H, Stadler P F and Bedau M A 2004 Transitions from nonliving to living matter *Science* **303** 963–5
[296] Bachmann P A, Luisi P L and Lang J 1992 Autocatalytic self-replicating micelles as models for prebiotic structures *Nature* **357** 57
[297] Walde P, Wick R, Fresta M, Mangone A and Luisi P L 1994 Autopoietic self-reproduction of fatty acid vesicles *JACS* **116** 11649–54

[298] Čejková J, Novak M, Štěpánek F and Hanczyc M M 2014 Dynamics of chemotactic droplets in salt concentration gradients *Langmuir* **30** 11937–44
[299] Hanczyc M M, Toyota T, Ikegami T, Packard N and Sugawara T 2007 Fatty acid chemistry at the oil- water interface: self-propelled oil droplets *J. Am. Chem. Soc.* **129** 9386–91
[300] Čejková J, Hanczyc M M and Štěpánek F 2018 Multi-armed droplets as shape-changing protocells *Artif. Life* **24** 71–9
[301] Čejkova J, Stepanek F and Hanczyc M M 2016 Evaporation-induced pattern formation of decanol droplets *Langmuir* **32** 4800–5
[302] Jitka Č, Schwarzenberger K, Eckert K and Tanaka S 2019 Dancing performance of organic droplets in aqueous surfactant solutions *Colloids Surf.,* A **566** 141–7
[303] Qiao Y, Li M, Booth R and Mann S 2017 Predatory behaviour in synthetic protocell communities *Nat. Chem.* **9** 110–9
[304] Čejková J, Banno T, Hanczyc M M and Štěpánek F 2017 Droplets as liquid robots *Artif. Life* **23** 528–49
[305] Kriegman S, Blackiston D, Levin M and Bongard J 2020 A scalable pipeline for designing reconfigurable organisms *Proc. Natl Acad. Sci.* **117** 1853–9
[306] Blackiston D, Lederer E, Kriegman S, Garnier S, Bongard J and Levin M 2021 A cellular platform for the development of synthetic living machines *Sci. Robot.* **6**
[307] Taylor T and Dorin A 2020 *Rise of the Self-Replicators* (Berlin: Springer)
[308] Bedau M A, McCaskill J S, Packard N H and Rasmussen S 2009 Living technology: exploiting life's principles in technology *Artif. Life* **16** 89–97 2011/11/03
[309] Bedau M A, McCaskill J S, Packard N H, Parke E C, S R and Rasmussen 2013 Introduction to recent developments in living technology *Artif. Life* **19** 291–8 2014/07/08
[310] Standish R K 2003 Open-ended artificial evolution *Int. J. Comput. Intell. Appl.* **3** 167–75
[311] Moreno A and Ruiz-Mirazo K 2006 The maintenance and open-ended growth of complexity in nature: information as a decoupling mechanism in the origins of life *Reframing Complexity: Perspectives from the North and South* ed F Capra, A Juarrero, P Sotolongo and J van Uden (ISCE Publishing)
[312] Taylor T, Bedau M, Channon A, Ackley D, Banzhaf W, Beslon G, Dolson E, Froese T, Hickinbotham S and Ikegami T *et al* 2016 Open-ended evolution: perspectives from the oee workshop in york *Artif. Life* **22** 408–23
[313] Roli A and Kauffman S A 2020 Emergence of organisms *Entropy* **22**
[314] Hernández-Orozco S, Hernández-Quiroz F and Zenil H 2018 Undecidability and irreducibility conditions for open-ended evolution and emergence *Artif. Life* **24** 56–70
[315] Kurt G 1931 Über formal unentscheidbare sätze der principia mathematica *u*nd verwandter systeme i *Monatsh. Math. Phys.* **38** 173–98
[316] Turing A M 1936 On computable numbers, with an application to the Entscheidungsproblem *Proc. London Math. Soc., Ser.* **2** 230–65
[317] Chaitin G J 1975 Randomness and mathematical proof *Sci. Am.* **232** 47–52
[318] Rahwan I *et al* 2019 Machine behaviour *Nature* **568** 477–86
[319] Nicolis G and Prigogine I 1977 *Self-Organization in Non-Equilibrium Systems: From Dissipative Structures to Order Through Fluctuations* (Chichester: Wiley)
[320] Martyushev L M and Seleznev V D 2006 Maximum entropy production principle in physics, chemistry and biology *Phys. Rep.* **426** 1–45
[321] Ramaswamy S 2017 Active matter *J. Stat. Mech: Theory Exp.* **2017** 054002

[322] van Leeuwenhoek A 1700 IV. Part of a letter from Mr Antony van Leeuwenhoek, concerning the worms in sheeps livers, gnats, and animalcula in the excrements of frogs *Phil. Trans. R. Soc.* **22** 509–18
[323] Corti B 1774 *Osservazioni Microscopische sulla Tremella e sulla Circolazione del Fluido in Una Pianta Acquajuola* (Italy: Appresso Giuseppe Rocchi, Lucca)
[324] Brown R 1828 A brief account of microscopical observations made in the months of June, July, and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies *Phil. Mag.* **4** 161–73
[325] Engelmann R W 1882 Ueber sauerstoffausscheidung van pfalnzenzellen im mikrospektrum *Pflügers Arch. Gesamte Physiol. Menschen Tiere* **27** 485–73
[326] Secchi E, Rusconi R, Buzzaccaro S, Salek M M, Smriga S, Piazza R and Stocker R 2016 Intermittent turbulence in flowing bacterial suspensions *J. R. Soc. Interface* **13** 20160175
[327] Haas P A, Höhn S, Honerkamp-Smith A R, Kirkegaard J B and Goldstein R E 2018 The noisy basis of morphogenesis: mechanisms and mechanics of cells sheet folding inferred from developmental variability *PLoS Biol.* **16** e2005536
[328] Besson S and Dumais J 2011 Universal rule for the symmetric division of plant cells *Proc. Natl. Acad. Sci. USA* **108** 6294–9
[329] Ratcliff W C, Denison R F, Borrello M and Travisano M 2012 Experimental evolution of multicellularity *Proc. Natl Acad. Sci. USA* **109** 1595–600
[330] Brunet T, Larsen B T, Linden T A, Vermeij M J A, McDonald K and King N 2019 Light-regulated collective contractility in a multicellular choanoflagellate *Science* **366** 326–34
[331] ESA 2020 Space for the oceans http://www.esa.int/Enabling_Support/Preparing_for_the_Future/Space_for_Earth/Space_for_the_oceans
[332] Hodgkin A L and Huxley A F 1952 A quantitative description of membrane current and its application to conductance and excitation in nerve *J. Physiol.* **117** 500–44
[333] Franklin R E and Goslin R G 1953 Molecular configuration in sodium thymonucleate *Nature* **171** 740–1
[334] Watson J D and Crick F H 1953 Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid *Nature* **171** 737–8
[335] Cole K S 1949 Dynamic electrical characteristics of the squid axon membrane *Arch. Sci. Physiol.* **3** 253–8
[336] Meron E 1992 Pattern formation in excitable media *Phys. Rep.* **218** 1–66
[337] Bragg W L 1913 The diffraction of short electromagnetic waves by a crystal *Proc. Camb. Phil. Soc.* **17** 43–57
[338] Murray J D 1989 *Mathematical Biology* (Berlin: Springer)
[339] Rayleigh L 1916 On convection currents in a horizontal layer of fluid, when the higher temperature is on the under side *Phil. Mag.* **32** 529–46
[340] Keller E F 2003 *Making Sense of Life. Explaining Biological Development with Models, Metaphors, and Machines* (Cambridge, MA: Harvard University Press)
[341] Castets V, Dulos E, Boissonade J and De Kepper P 1990 Experimental evidence of a sustained standing turing-type nonequilibrium chemical pattern *Phys. Rev. Lett.* **64** 2953–6
[342] De Kepper P, Castets V, Dulos E and Boissonade J 1991 Turing-type chemical patterns in the chlorite-iodide-malonic acid reaction *Physica* D **49** 161–9
[343] Ouyang Q and Swinney H L 1991 Transition from a uniform state to hexagonal and striped turing patterns *Nature* **352** 610–2

[344] Goldstein R E 2018 Coffee stains, cell receptors, and time crystals: lessons from the old literature *Phys. Today* **71** 32–8
[345] Keller E F and Segel L A 1971 Model for chemotaxis *J. Theor. Biol.* **30** 225–34
[346] Lee K J, Cox E C and Goldstein R E 1996 Competing patterns of signaling activity in *dictyostelium discoideum Phys. Rev. Lett.* **76** 1174–7
[347] Taylor G I 1951 Analysis of the swimming of microscopic organism. *Proc. R. Soc.* A **209** 447–61
[348] Lighthill M J 1952 On the squirming motion of nearly spherical deformable bodies through liquids at very small Reynolds numbers *Commun. Pure Appl. Math.* **5** 109–18
[349] Purcell E M 1977 Life at low Reynolds number *Am. J. Phys.* 45–11
[350] Berg H C and Purcell E M 1977 Physics of chemoreception *Biophys. J.* **20** 193–219
[351] Steinberg M S 1962 On the mechanism of tissue reconstruction by dissociated cells, I. Population kinetics, differential adhesiveness, and the absence of directed migration *Proc. Natl Acad. Sci. USA* **48** 1577–82
[352] Foty R A, Forgacs G, Pfleger C M and Steinberg M S 1994 Liquid properties of embryonic tissues: measurement of interfacial tensions *Phys. Rev. Lett.* **72** 2298–301
[353] de Gennes P G 1979 *Scaling Concepts in Polymer Physics* (Ithaca, NY: Cornell University Press)
[354] Flory P J 1953 *Principles of Polymer Chemistry* (Ithaca, NY: Cornell University Press)
[355] Langer J S 1980 Instabilities and pattern formation in crystal growth *Rev. Mod. Phys.* **52** 1–28
[356] Helfrich W 1973 Elastic properties of lipid bilayers: theory and possible experiments *Z. Naturforsch.* C **28** 693–703
[357] Smith S B, Finzi L and Bustamante C 1992 Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads *Science* **258** 1122–6
[358] Marko J F and Siggia E D 1995 Stretching DNA *Macromolecules* **28** 8759–70
[359] Svoboda K, Schmidt C F, Schnapp B J and Block S M 1993 Direct observation of kinesin stepping by optical trapping interferometry *Nature* **365** 721–7
[360] Magnasco M O 1993 Forced thermal ratchets *Phys. Rev. Lett.* **71** 1477–80
[361] Rousselet J, Salome L, Ajdari A and Prost J 1994 Directional motion of Brownian particles induced by a periodic asymmetric potential *Nature* **370** 446–7
[362] Jülicher F, Ajdari A and Prost J 1997 Modeling molecular motors *Rev. Mod. Phys.* **69** 1269–81
[363] Bustamante C, Keller D and Oster G 2001 The physics of molecular motors *Acc. Chem. Res.* **34** 412–20
[364] Lubkins S and Rand R 1994 Oscillatory reaction-diffusion equations on rings *J. Math. Biol.* **32** 617–32
[365] Hänggi P and Marchesoni F 2009 Artificial Brownian motors: controlling transport on the nanoscale *Rev. Mod. Phys.* **81** 387–442
[366] Guerin T, Prost J, Martin P and Joanny J F 2010 Coordination and collective properties of molecular motors: theory *Curr. Opin. Cell Biol.* **22** 14–20
[367] Nadrowski B, Martin P and Jülicher F 2004 Active hair-bundle motility harnesses noise to operate near an optimum of mechanosensitivity *Proc. Natl Acad. Sci. USA* **101** 12195–200
[368] Crocker J C and Grier D G 1996 Methods of digital video microscopy for colloidal studies *J. Coll. Int. Sci.* **179** 298–310

[369] Polin M, Tuval I, Drescher K, Gollub J P and Goldstein R E 2009 *Chlamydomonas* swis with two 'gears' in a eukaryotic version of run-and-tumble locomotion *Science* **325** 487–90

[370] Wan K Y and Goldstein R E 2018 Time-irreversibility and criticality in the motility of a flagellate microorganism *Phys. Rev. Lett.* **121** 058103

[371] Squires T M and Quake S R 2005 Microfluidics: fluid physics at the nanoliter scale *Rev. Mod. Phys.* **77** 977–1026

[372] Huebner J S, Jakobsson E, Dam H G, Parsegian V A and Austin R H 1997 How should physicists, biologists work together? The 'Harness the hubris' debate continues *Phys. Today* **50** 11–95

[373] Parsegian V A 1997 Harness the hubris: useful things physicists could do in biology *Phys. Today* **50** 23–7

[374] Finlayson B A and Scriven L E 1969 Convective instability by active stress *Proc. R. Soc.* A **310** 183–219

[375] Marchetti M C, Joanny J F, Ramaswamy S, Liverpool T B, Prost J, Rao M and Simha R A 2013 Hydrodynamics of soft active matter *Rev. Mod. Phys.* **85** 1143–89

[376] Ramaswamy S 2019 Active fluids *Nat. Rev. Phys.* **1** 640–64

[377] Gompper G, Winkler R G, Speck T, Solon A, Nardini C, Peruani F, Löwen H, Golestanian R, Kaupp U B and Alvarez L 2020 The 2020 motile active matter roadmap *J. Phys. Condens. Matter* **32** 193001

[378] Shaebani M R, Wysocki A, Winkler R G, Gompper G and Rieger H 2020 Computational models for active matter *Nat. Rev. Phys.* **2** 181–99

[379] Prost J and Bruinsma R 1996 Shape fluctuations of active membranes *Europhys. Lett.* **33** 321–6

[380] Ramaswamy S, Prost J and Toner J 2000 Nonequilibrium fluctuations, travelling waves, and instabilities in active membranes *Phys. Rev. Lett.* **84** 3494–7

[381] Nédélec F J, Surrey T, Maggs A C and Leibler S 1997 Self-organization of microtubules and motors *Nature* **389** 305–8

[382] Toner J and Tu Y 1995 Long-range order in a two-dimensional dynamical XY model: how birds fly together *Phys. Rev. Lett.* **75** 4326–9

[383] Simha R A and Ramaswamy S 2002 Hydrodynamic fluctuations and instabilities in ordered suspensions of self-propelled particles *Phys. Rev. Lett.* **89** 058101

[384] Saintillan D and Shelley M J 2008 Instabilities and pattern formation in active particle suspensions: kinetic theory and continuum simulations *Phys. Rev. Lett.* **100** 178103

[385] Dombrowski C, Cisneros L, Chatkaew S, Kessler J O and Goldstein R E 2004 Self-concentration and large-scale coherence in bacterial dynamics *Phys. Rev. Lett.* **93** 098103

[386] Lacoste D and Bassereau P 2014 An update on active membranes ed G Pabst, N KuÄerka, M-P Nieh and J Katsaras *Liposomes, Lipid Bilayers and Model Membranes: From Basic Research to Application* pp 271–87 (Boca Raton, FL: CRC Press, Taylor and Francis Group)

[387] Turlier H and Betz T 2019 Unveiling the active nature of living-membrane fluctuations and mechanics *Annu. Rev. Condens. Matter Phys.* **10** 213–32

[388] Turlier H, Fedosov D A, Audoly B, Auth T, Gov N S, Sykes C, Joanny J-F, Gompper G and Betz T 2016 Equilibrium physics breakdown reveals the active nature of red blood cell flickering *Nat. Phys.* **12** 513–9

[389] Kruse K, Joanny J F, Jülicher F, Prost J and Sekimoto K 2005 Generic theory of active polar gels: a paradigm for cytoskeletal dynamics *Eur. Phys. J.* E **16** 5–16

[390] Prost J, Jülicher F and Joanny J F 2015 Active gel physics *Nat. Phys.* **11** 111–7

[391] Doostmohammadi A, Ignés-Mullol J, Yeomans J M and Sagués F 2018 Active nematics *Nat. Commun.* **9** 3246

[392] Stenhammar J, Wittkowski R, Marenduzzo D and Cates M E 2015 Activity-induced phase separation and self-assembly in mixtures of active and passive particles *Phys. Rev. Lett.* **114** 018301

[393] Banani S F, Lee H O, Hyman A A and Rosen M K 2017 Biomolecular condensates: organizers of cellular biochemistry *Nat. Rev. Mol. Cell Biol.* **18** 285–98

[394] Shin Y, Chang Y-C, Lee D S W, Berry J, Sanders D W, Ronceray P, Wingreen N S, Haataja M and Brangwynne C P 2018 Liquid nuclear condensates mechanically sense and restructure the genome *Cell* **175** 1481–91

[395] Zwicker D, Hyman A A and Jülicher F 2015 Suppression of Ostwald ripening in active emulsions *Phys. Rev.* E **92** 012317

[396] Guillot C and Lecuit T 2013 Mechanics of epithelial tissue homeostasis and morphogenesis *Science* **340** 1185–9

[397] Collinet C, Rauzi M, Lenne P-F and Lecuit T 2015 Local and tissue-scale forces drive oriented junction growth during tissue extension *Nat. Cell Biol.* **17** 1247–58

[398] Ranft J, Basan M, Elgeti J, Joanny J-F, Prost J and Jülicher F 2010 Fluidization of tissues by cell division and apoptosis *Proc. Natl Acad. Sci. USA* **107** 20863–8

[399] PopoviÄ M, Nandi A, Merkel M, Etournay R, Eaton S, Jülicher F and Salbreux G 2017 Active dynamics of tissue shear flow *New J. Phys.* **19** 033006

[400] Girkin J M and Carvalho M T 2018 The light-sheet microscopy revolution *J. Opt.* **20** 053002

[401] Takebe T and Wells J M 2019 Organoids by design *Science* **364** 956–9

[402] Haas P A and Goldstein R E 2019 Nonlinear and nonlocal elasticity in coarse-grained differential-tension models of epithelia *Phys. Rev.* E **99** 022411

[403] Vagne Q, Vrel J-P and Sens P 2020 A minimal self-organisation model of the Golgi apparatus *eLife* **9** e47318

[404] Thompson D W 1942 *On Growth and Form* 2nd edn (Cambridge: Cambridge University Press)

[405] Discher D E, Janmey P and Wang Y-L 2005 Tissue cells feel and respond to the stiffness of their substrate *Science* **310** 1139–43

[406] Harris A R, Jreij P and Fletcher D A 2016 Mechanotransduction by the actin cytoskeleton: converting mechanical stimuli into biochemical signals *Annu. Rev. Biophys.* **47** 617–31

[407] Engler A J, Sen S, Sweeney H L and Discher D E 2006 Matrix elasticity directs stem cell lineage specification *Cell* **126** 677–89

[408] Fernandez-Sanchez M-E, Brunet T, Röper J-C and Farge E 2015 Mechanotransduction's impact on animal development, evolution, and tumorigenesis *Annu. Rev. Cell Dev. Biol.* **31** 373–97

[409] Broders-Bondon F, Nguyen Ho-Bouldoires T H, Fernandez-Sanchez M-E and Farge E 2018 Mechanotransduction in tumor progression: the dark side of the force *J. Cell Biol.* **217** 1571–87

[410] Roca-Cusachs P, Conte V and Trepat X 2017 Quantifying forces in cell biology *Nat. Cell Biol.* **19** 742–51

[411] Geiger B, Bershadsky A, Pankov R and Yamada K M 2001 Transmembrane crosstalk between the extracellular matrix and the cytoskeleton *Nat. Rev. Mol. Cell Biol.* **2** 793–805

[412] Martino F, Perestrelo A R, Vinarský V, Pagliari S and Forte G 2018 Cellular mechanotransduction: from tension to function *Front. Physiol.* **9** 824

[413] Charras G and Yap A S 2018 Tensile forces and mechanotransduction at cell–cell junctions *Curr. Biol.* **28** R445–57

[414] Gottlieb P A (ed) 2017 *Piezo Channels, Volume 79 of Current Topics in Membranes* (New York: Academic)

[415] Eder D, Aegerter C and Basler K 2017 Forces controlling organ growth and size *Mech. Develop.* **144** 53–61

[416] Van Essen D C, Smith S M, Barch D M, Behrens T E J, Yacoub E and Ugurbil K 2013 The WU-Minn human connectome project: an overview *NeuroImage* **80** 62–79

[417] Zecchi K A and Heimburg T 2021 Non-linear conductance, rectification, and mechanosensitive channel formation of lipid membranes *Front. Cell Dev. Biol.* **8** 1751

[418] Deubner J, Coulon P and Diester I 2019 Optogenetic approaches to study the mammalian brain *Curr. Opin. Struct. Biol.* **57** 157–63

[419] Calarco J A and Samuel A D T 2019 Imaging whole nervous systems: insights into behavior from brains inside bodies *Nat. Meth.* **16** 14–5

[420] Dragomir E I, Åtih V and Portugues R 2020 Evidence accumulation during a sensorimotor decision task revealed by whole-brain imaging *Nat. Neurosci.* **23** 85–93

[421] Gao R *et al* 2019 Cortical column and whole-brain imaging with molecular contrast and nanoscale resolution *Science* **363** 543–8

[422] Richards B A *et al* 2019 A deep learning framework for neuroscience *Nat. Neurosci.* **22** 1761–70

[423] Bassett D S and Sporns O 2017 Network neuroscience *Nat. Neurosci.* **20** 353–64

[424] Schrödinger E 1944 *What is Life? The Physical Aspect of the Living Cell* (Cambridge: Cambridge University Press)

[425] Xu C, Hu S and Chen X 2016 Artificial cells: from basic science to applications *Mat. Today* **19** 516–32

[426] https://nature.com/collections/ngmpdftlqt/

[427] Zwicker D, Seyboldt R, Weber C A, Hyman A A and Jülicher F 2017 Growth and division of active droplets provides a model for protocells *Nat. Phys.* **13** 408–13

[428] Jia H and Schwille P 2019 Bottom-up synthetic biology: reconstitution in space and time *Curr. Opin. Biotech.* **60** 179–87

[429] Gibson D G *et al* 2010 Creation of a bacterial cell controlled by a chemically synthesized genome *Science* **329** 52–6

[430] Breuer M *et al* 2019 Essential metabolism for a minimal cell *eLife* **8** e36842

[431] Braun D, Goddard N L and Libchaber A 2003 Exponential DNA replication by laminar convection *Phys. Rev. Lett.* **91** 158103

[432] Kreysing M, Keil L, Lanzmich S and Braun D 2015 Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length *Nat. Chem.* **7** 203–8

[433] Bell G and Mooers A O 1997 Size and complexity among multicellular organisms *Biol. J. Linn. Soc.* **60** 345–63

[434] Weismann A 1892 *Essays on Heredity and Kindred Biological Problems* (Oxford: Oxford University Press)

[435] Huxley J S 1912 *The Individual in the Animal Kingdom* (Cambridge: Cambridge University Press)

[436] Solari C A, Kessler J O and Michod R E 2006 A hydrodynamics approach to the evolution of multicellularity: flagellar motility and germ-soma differentiation in volvocalean green algae *Am. Nat.* **167** 537–54
[437] Goldstein R E 2015 Green algae as model organisms for biological fluid dynamics *Annu. Rev. Fluid Mech.* **47** 343–75
[438] Brunet T and King N 2017 The origin of animal multicellularity and cell differentiation *Dev. Cell* **43** 124–40
[439] Sella G and Hirsh A E 2005 The application of statistical physics to evolutionary biology *Proc. Natl Acad. Sci. USA* **102** 9541
[440] Drossel B 2001 Biological evolution and statistical physics *Adv. Phys.* **50** 209–95
[441] Wolf Y I, Katsnelson M I and Koonin E V 2018 Physical foundations of biological complexity *Proc. Natl Acad. Sci. USA* **115** E8678
[442] Kasting J F 1993 Earth's early atmosphere *Science* **259** 920–6
[443] Kasting J F, Toon O B and Pollack J B 1988 How climate evolved on the terrestrial planets *Sci. Am.* **258** 90–7
[444] Bottke W F, Vokrouhlický D, Marchi S, Swindle T, Scott E R D, Weirich J R and Levison H 2015 Dating the Moon-forming impact event with asteroidal meteorites *Science* **348** 321–3
[445] Schlichting H E and Mukhopadhyay S 2018 Atmosphere impact losses *Space Sci. Rev.* **214** 34
[446] Lebrun T, Massol H, ChassefièRe E, Davaille A, Marcq E, Sarda P, Leblanc F and Brandeis G 2013 Thermal evolution of an early magma ocean in interaction with the atmosphere *J. Geophys. Res. (Planets)* **118** 1155–76
[447] Stüeken E E, Som S M and Claire M *et al* 2020 Mission to planet Earth: the first two billion years *Space Sci. Rev.* **216** 31
[448] Pahlevan K, Schaefer L and Hirschmann M M 2019 Hydrogen isotopic evidence for early oxidation of silicate Earth *Earth Planet. Sci. Lett.* **526** 115770
[449] Sleep N H 2010 The Hadean-Archaean environment *Cold Spring Harb. Perspect. Biol.* **2** a002527
[450] Bekaert D V, Broadley M W and Marty B 2020 The origin and fate of volatile elements on Earth revisited in light of noble gas data obtained from comet 67P/Churyumov–Gerasimenko *Sci. Rep.* **10** 5796
[451] Mezger K, Maltese A and Vollstaedt H 2021 Accretion and differentiation of early planetary bodies as recorded in the composition of the silicate earth *Icarus* **365** 114497
[452] Mojzsis S J, Brasser R, Kelly N M, Abramov O and Werner S C 2019 Onset of Giant planet migration before 4480 million years ago *Astrophys. J.* **881** 44
[453] Fuller-Rowell T, Solomon S, Roble R and Viereck R 2004 Impact of solar EUV, XUV, and x-ray variations on Earth's atmosphere *Solar Variability and its Effects on Climate. Geophysical Monograph 141* (AGU Publications) vol 141 p 341
[454] Lammer H, Leitzinger M and Scherf M *et al* 2020 Constraining the early evolution of Venus and Earth through atmospheric Ar, Ne isotope and bulk K/U ratios *Icarus* **339** 113551
[455] Ribas I, Guinan E F, Güdel M and Audard M 2005 Evolution of the solar activity over time and effects on planetary atmospheres. I. High-energy irradiances (1–1700 Å) *Astrophys. J.* **622** 680–94
[456] Sagan C and Mullen G 1972 Earth and Mars: evolution of atmospheres and surface temperatures *Science* **177** 52–6

[457] Catling D C and Kasting J F 2007 *Planetary Atmospheres and Life* (London: Cambridge University Press)
[458] Elkins-Tanton L T 2008 Linked magma ocean solidification and atmospheric growth for Earth and Mars *Earth Planet. Sci. Lett.* **271** 181–91
[459] Guzmán-Marmolejo A, Segura A and Escobar-Briones E 2013 Abiotic production of methane in terrestrial planets full access *Astrobiology* **13** 550–9
[460] Atreya S K, Adams E Y, Niemann H B, Demick-Montelara J E, Owen T C, Fulchignoni M, Ferri F and Wilson E H 2006 Titan's methane cycle *Planet. Space Sci.* **54** 1177–87
[461] Pavlov A A, Brown L L and Kasting J F 2001 UV shielding of NH_3 and O_2 by organic hazes in the Archean atmosphere *J. Geophys. Res.* **106** 23267–88
[462] Heller R, Duda J-P, Winckler M, Reitner J and Gizon L 2021 Habitability of the early Earth: liquid water under a faint young Sun facilitated by strong tidal heating due to a nearby Moon *Palaontol. Z.* **95** 563–75
[463] Pope E C, Bird D K and Rosing M T 2012 Isotope composition and volume of Earth's early oceans *Proc. Natl Acad. Sci.* **109** 4371–6
[464] Sharp Z D, McCubbin F M and Shearer C K 2013 A hydrogen-based oxidation mechanism relevant to planetary formation *Earth Planet. Sci. Lett.* **380** 88–97
[465] Marty B, Avice G, Bekaert D V and Broadley M W 2018 Salinity of the Archaean oceans from analysis of fluid inclusions in quartz *C.R. Geosci.* **350** 154–63
[466] Matthes K, Funke B and Andersson M E *et al* 2017 Solar forcing for CMIP6 (v3.2) *Geosci. Model Dev.* **10** 2247–302
[467] Pascal R, Pross A and Sutherland J D 2013 Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics *Open Biol* **3** 130156
[468] James Cleaves H, Chalmers J H, Lazcano A, Miller S L and Bada J L 2008 A reassessment of prebiotic organic synthesis in neutral planetary atmospheres *Origins Life Evol. Biosphere* **38** 105–15
[469] Martin W F, Baross J, Kelley D and Russell M J 2008 Hydrothermal vents and the origin of life *Nat. Rev. Microbiol.* **6** 805–14
[470] Shock E L and Schulte M D 1998 Organic synthesis during fluid mixing in hydrothermal systems *J. Geophys. Res.* **103** 28513–28
[471] Bertrand M, Chabin A, Brack A, Cottin H, Chaput D and Westall F 2012 The PROCESS experiment: exposure of amino acids in the EXPOSE—experiment on the International Space Station and in laboratory simulations *Astrobiology* **12** 426–35
[472] Marty B, Alexander C M O and Raymond S N 2013 Primordial Origins of Earth's Carbon *Rev. Mineral. Geochem.* **75** 149–181
[473] Duprat J, Dobrică E and Engrand C *et al* 2010 Extreme deuterium excesses in ultracarbonaceous micrometeorites from central antarctic snow *Science* **328** 742
[474] Schmitt-Kopplin P, Gabelica Z and Gougeon R D *et al* 2010 High molecular diversity of extraterrestrial organic matter in Murchison meteorite revealed 40 years after its fall *Proc. Natl Acad. Sci.* **107** 2763–8
[475] Lowe D R, Byerly G R and Kyte F T *et al* 2003 Spherule beds 3.47–3.24 billion years old in the Barberton Greenstone Belt, South Africa: a record of large meteorite impacts and their influence on early crustal and biological evolution *Astrobiology* **3** 7–48
[476] Lowe D R, Byerly G R and Kyte F T 2014 Recently discovered 3.42–3.23 Ga impact layers, Barberton Belt, South Africa: 3.8 Ga detrital zircons, Archean impact history, and tectonic implications *Geology* **42** 747–50

[477] Gourier D, Binet L and Calligaro T *et al* 2019 Extraterrestrial organic matter preserved in 3.33 Ga sediments from Barberton, South Africa *Geochim. Cosmochim. Acta* **258** 207–25

[478] Sasselov D D, Grotzinger J P and Sutherland J D 2020 The origin of life as a planetary phenomenon *Sci. Adv.* **6** eaax3419

[479] Pascal R 2012 Suitable energetic conditions for dynamic chemical complexity and the living state *J. Syst. Chem.* **3**

[480] Baross J A and Hoffman S E 1985 Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life *Orig. Life* **15** 327–45 :

[481] Russell M J and Hall A J 1997 The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front *J. Geol. Soc.* **154** 377–402

[482] Russell M J, Hall A J and Martin W 2010 Serpentinization as a source of energy at the origin of life *Geobiology* **8** 355–71

[483] Westall F, Hickman-Lewis K and Hinman N *et al* 2018 A hydrothermal-sedimentary context for the origin of life *Astrobiology* **18** 259–93

[484] Cottin H, Guan Y Y and Noblet A *et al* 2012 The PROCESS experiment: an astrochemistry laboratory for solid and gaseous organic samples in low-earth orbit *Astrobiology* **12** 412–25

[485] Deamer D W, Singaram S, Rajamani S, Kompanichenko V and Guggenheim S 2006 Self-assembly processes in the prebiotic environment *Phil. Trans. R. Soc.* B **361** 1809–18

[486] Tartese R, Chaussidon M, Gurenko A, Delarue F and Robert F 2017 Archean oceans reconstructed from oxygen isotope composition of early-life remnants *Geochem. Persp. Lett.* **3** 55 65

[487] Usoskin I G 2017 A history of solar activity over millennia *Living Rev. Sol. Phys.* **14** 3

[488] Badalyan O G and Obridko V N 2011 North-south asymmetry of the sunspot indices and its quasi-biennial oscillations *Nature* **16** 357–65

[489] Kane R P 2005 Differences in the quasi-biennial oscillation and quasi-triennial oscillation characteristics of the solar, interplanetary, and terrestrial parameters *J. Geophys. Res. (Space Phys.)* **110** A01108

[490] Broomhall A-M, Fletcher S T and Salabert D *et al* 2011 Are short-term variations in solar oscillation frequencies the signature of a second solar dynamo? *J. Phys.: Conf. Ser.* **271** 012025 (GONG-SoHO 24: A New Era of Seismology of the Sun and Solar-Like Stars)

[491] Manuel G 2007 The Sun in time: activity and environment *Living Rev. Sol. Phys.* **4** 3

[492] Claire M W, Sheets J, Cohen M, Ribas I, Meadows V S and Catling D C 2012 The evolution of solar flux from 0.1 nm to 160 μm: quantitative estimates for planetary studies *Astrophys. J.* **757** 95

[493] Bahcall J N, Pinsonneault M H and Basu S 2001 Solar models: current epoch and time dependences, neutrinos, and helioseismological properties *Astrophys. J.* **555** 990–1012

[494] Lean J L, Rottman G J, Lee Kyle H, Woods T N, Hickey J R and Puga L C 1997 Detection and parameterization of variations in solar mid- and near- ultraviolet radiation (200–400 nm) *J. Geophys. Res.* **102** 29939–56

[495] Soderblom D R, Stauffer J R, MacGregor K B and Jones B F 1993 The evolution of angular momentum among zero-age main-sequence solar-type stars *Astrophys. J.* **409** 624

[496] David Dorren J and Guinan E F 1994 HD 129333: the Sun in its infancy *Astrophys. J.* **428** 805

[497] Brun A S and Browning M K 2017 Magnetism, dynamo action and the solar-stellar connection *Living Rev. Sol. Phys.* **14** 4

[498] Canuto V M, Levine J S, Augustsson T R and Imhoff C L 1982 UV radiation from the young Sun and oxygen and ozone levels in the prebiological palaeoatmosphere *Nature* **296** 816–20
[499] Meert J G, Levashova N M, Bazhenov M L and Landing E 2016 Rapid changes of magnetic field polarity in the late Ediacaran: linking the Cambrian evolutionary radiation and increased UV-B radiation *Gondwana Res.* **34** 149–57
[500] Airapetian V S, Glocer A, Gronoff G, Hébrard E and Danchi W 2016 Prebiotic chemistry and atmospheric warming of early Earth by an active young Sun *Nat. Geosci.* **9** 452–5
[501] Austin J, Tourpali K and Rozanov E *et al* 2008 Coupled chemistry climate model simulations of the solar cycle in ozone and temperature *J. Geophys. Res. (Atmos.)* **113** D11306
[502] Gray L J, Beer J, Geller M and Haigh J D *et al* 2010 Solar influences on climate *Rev. Geophys.* **48** RG4001
[503] Haigh J D 1996 The impact of solar variability on climate *Science* **272** 981–4
[504] Matthes K, Kuroda Y, Kodera K and Langematz U 2006 Transfer of the solar signal from the stratosphere to the troposphere: northern winter *J. Geophys. Res.* **111** D06108
[505] Matthes K, Biastoch A and Wahl S *et al* 2020 The flexible ocean and climate infrastructure version 1 (FOCI1): mean state and variability *Geosci. Model Dev.* **13** 2533–68
[506] Sinnhuber M, Nieder H and Wieters N 2012 Energetic particle precipitation and the chemistry of the mesosphere/lower thermosphere *Surv. Geophys.* **33** 1281–334
[507] Dunne E M, Gordon H and Kürten A *et al* 2016 Global atmospheric particle formation from CERN CLOUD measurements *Science* **354** 1119–24
[508] Mojzsis S J, Harrison T M and Pidgeon R T 2001 Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago *Nature* **409** 178–81
[509] Wilde S A, Valley J W, Peck W H and Graham C M 2001 Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago *Nature* **409** 175–8
[510] Whitehouse M J, Nemchin A A and Pidgeon R T 2017 What can Hadean detrital zircon really tell us? A critical evaluation of their geochronology with implications for the interpretation of oxygen and hafnium isotopes *Gondwana Res.* **51** 78 91
[511] Arndt N T and Nisbet E G 2012 Processes on the young earth and the habitats of early life *Annu. Rev. Earth Planet. Sci.* **40** 521–49
[512] Kamber B S 2015 The evolving nature of terrestrial crust from the Hadean, through the Archaean, into the Proterozoic *Precambrian Res.* **258** 48 82
[513] Wang Z and Becker H 2013 Ratios of S, Se and Te in the silicate Earth require a volatile-rich late veneer *Nature* **499** 328–31
[514] Morbidelli A and Wood B J 2015 *Late Accretion and the Late Veneer* (Washington, DC: American Geophysical Union (AGU))
[515] Biggin A J, de Wit M J, Langereis C G, Zegers T E, Voûte S, Dekkers M J and Drost K 2011 Palaeomagnetism of Archaean rocks of the Onverwacht Group, Barberton Greenstone Belt (southern Africa): evidence for a stable and potentially reversing geomagnetic field at ca. 3.5 Ga *Earth Planet. Sci. Lett.* **302** 314–28
[516] Usui Y, Tarduno J A, Watkeys M, Hofmann A and Cottrell R D 2009 Evidence for a 3.45-billion-year-old magnetic remanence: hints of an ancient geodynamo from conglomerates of South Africa *Geochem. Geophys. Geosyst.* **10** Q09Z07

[517] Tarduno J A, Cottrell R D and Bono R K *et al* 2020 Paleomagnetism indicates that primary magnetite in zircon records a strong Hadean geodynamo *Proc. Natl Acad. Sci.* **117** 2309–18
[518] Borlina C S, Weiss B P and Lima E A *et al* 2020 Reevaluating the evidence for a Hadean-Eoarchean dynamo *Sci. Adv.* **6** eaav9634
[519] Stevenson D J 2001 Mars' core and magnetism *Nature* **412** 214–9
[520] Lockwood M, Harrison R G, Woollings T and Solanki S K 2010 Are cold winters in Europe associated with low solar activity? *Environ. Res. Lett.* **5** 024001