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 6

Physics for the environment and sustainable development

Ankit Agarwal, Philippe Azais, Luisa Cifarelli, Jacob de Boer, Deniz Eroglu, Gérard Gebel, Carlos Hidalgo, Didier Jamet, Juergen Kurths, Florence Lefebvre-Joud, Søren Linderoth, Alberto Loarte, Norbert Marwan, Natalio Mingo, Ugur Ozturk, Simon Perraud, Robert Pitz-Paal, Stefaan Poedts, Thierry Priem, Bernd Rech, Marco Ripani, Shubham Sharma, Tuan Quoc Tran and Hermann-Josef Wagner

6.1 Introduction

Luisa Cifarelli[1] and Carlos Hidalgo[2]
[1]University of Bologna, Bologna, Italy
[2]CIEMAT, Laboratorio Nacional de Fusión, Madrid, Spain

One of the most crucial and challenging developments of recent decades has been the discovery that the environment is fragile. This discovery shows that we cannot afford to delay the implementation of actions to tackle climate change if the long-term objective is to limit the increase in temperature of the planet at an affordable cost. Although the effects of climate change on the environment are too complex to admit simple solutions, recent developments illustrate how basic science can be pulled together successfully with social awareness and political action to avert an environmental tragedy.

This section presents work done in a wide range of research areas, illustrating how humanity has the responsibility to preserve our delicate planet but also the power to affect its environment. The chapters highlight the strength of fully interdisciplinary effort among physicists, mathematicians, and chemists as well as multilateral science to address global challenges that affect societies at their core.

doi:10.1088/978-0-7503-6342-6ch6

A thorough understanding of the Earth's system is essential for the life quality of modern society. It is important to be able to define the conditions for sustainable development of humanity in order to maintain the Earth's system within habitable limits, predict critical transitions and events in the Earth's dynamics, and effectively mitigate and adapt to changes and events related to the Earth's system to prevent the disastrous consequences of natural hazards. Section 6.2 deals with Earth system analysis from a nonlinear physics perspective. It describes key concepts from nonlinear physics and shows that they enable us to treat challenging problems of Earth sciences.

Energy is the lifeblood of today's society and one of the factors that has decisively contributed to improving humanity's quality of life. Section 6.3 deals with the description of physics fields with relevance for energy technologies. It addresses the further development of energy sources, such as solar, wind, nuclear fission energy, and storing energy storage systems as well as the quest to develop nuclear fusion, since the dominance of fossil fuels must decline. It addresses the potential challenges and opportunities in the development of global energy systems, emphasising how deeply interconnected the energy and climate debates are.

The invention of the internal combustion engine radically transformed industrial and personal transport and, consequently, our social organization system. Section 6.4 deals with transport electrification for green cities. It addresses research and development to deploy technologies that enhance the performance of electric drive vehicles.

Hazardous wastes and materials are diverse, with compositions and properties that vary significantly between industries and related energy sources. Section 6.5 deals with environmental safety from a chemical perspective to address how environmental emissions and waste disposal can be managed to meet sustainable development criteria.

Finally, space weather describes the way in which the Sun, through emergence of magnetic field into its atmosphere, flares, coronal mass emissions, high-energy particles, and subsequently induced space conditions, affects human activity and technology both in space and on the ground. Section 6.6 invites us to understand and predict space weather.

6.2 Earth system analysis from a nonlinear physics perspective

Juergen Kurths[1], Ankit Agarwal[1,2], Ugur Ozturk[3,4], Shubham Sharma[3], Norbert Marwan[1,5] and Deniz Eroglu[6]
[1]PIK—Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, 14473 Potsdam, Germany
[2]Department of Hydrology, Indian Institute of Technology Roorkee, 247667 Roorkee, India
[3]Helmholtz Centre Potsdam–GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany
[4]Institute of Environmental Science and Geography, University of Potsdam, 14476 Potsdam, Germany
[5]Institute of Geosciences, University of Potsdam, 14476 Potsdam, Germany
[6]Faculty of Engineering and Natural Sciences, Kadir Has University, 34083 Istanbul, Turkey

A reliable understanding of the Earth's system is essential for a good quality of life for modern society. Natural hazards are the cause of most life and resource losses. The ability to define the conditions for sustainable development of humankind, to keep the Earth's system within the boundaries of habitable states, and to predict critical transitions and events in the dynamics of the Earth's system are crucial to mitigate and adapt to Earth system–related events and changes (e.g., volcanic eruptions, earthquakes, and climate change) and to avert the disastrous consequences of natural hazards. In this chapter, we discuss key concepts from nonlinear physics and show that they enable us to treat challenging problems of Earth sciences which cannot be solved by classic methods. In particular, the concepts of multiscaling, recurrence, synchronization, and complex networks have become crucial in recent decades for a substantially more profound understanding of the dynamics of earthquakes, landslides, and (paleo)climate. They can even provide a significantly improved prediction of several high-impact extreme events. Additionally, crucial open challenges in the realm of methodological nature and applications to Earth sciences are given.

6.2.1 Introduction

The invention of thermoscopes and barometers in the early 17th century enabled the study of physical parameters of climate variables, such as precipitation, temperature, and pressure. Exploring the Earth's system in detached disciplinary practices became convenient with these early instruments for limited geographic locations. The disciplinary assessment of individual Earth system components continues to help in understanding fundamental mechanisms. They have been regarded as autonomous systems in their own right and further broken down into more specialized subsystems. One standard topic is to study, for instance, precipitation concerning more prominent atmospheric modes [1]. Until recent decades, this traditional practice of studying the four major spheres of the Earth's system, that is, the atmosphere, hydrosphere, biosphere, and geosphere, continued independently. However, the Earth behaves as an integrated complex system with nonlinear interactions and feedback loops between and within them [2]. For example, the influence of significant volcanic eruptions on climate oscillations proves a vital link between the geosphere and the atmosphere [3]. The increasing availability of data and the rising concerns related to shifts in the global climate system, concomitant extremes, and natural hazards have urged the development of a more *holistic understanding of the Earth's system* in recent decades (figure 6.1).

Furthermore, it was assumed that various Earth processes are *scale-invariant*, that is, we can expect a phenomenon to occur in several scales when we observe its occurrence on only one scale [4]. Indeed, the scale invariance theory was applied to many fields, such as the frequency–size distributions of rock fragments, faults, earthquakes, volcanic eruptions, landslides, and oil fields, but not all on the Earth's systems. Nevertheless, even the lack of scale invariance means that information is stored and perceived differently at different scales, resulting from mutual interactions of intertwined subcomponents interacting over a wide range of scales. Generally, a deep understanding of these multicomponent interactions between the different subsystems of the Earth's system, including human activities, requires an interdisciplinary approach in which concepts from various fields of physics and complex systems science are vital elements [5].

Trying to understand interacting Earth systems as a giant complex system using only instrumental records is insufficient, since such measures cover only a very

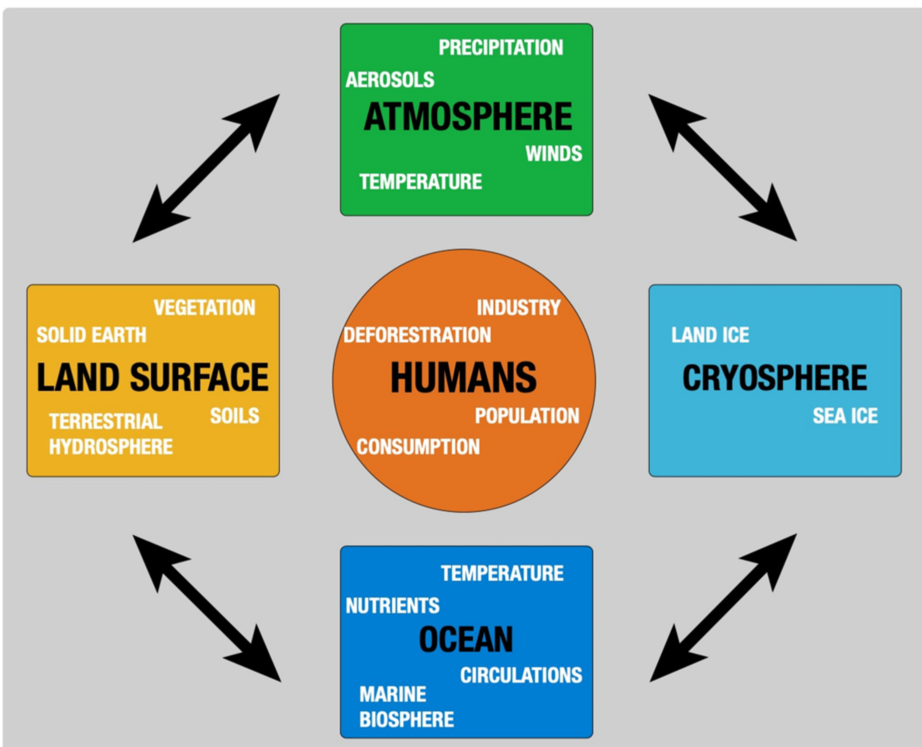


Figure 6.1. Scheme of rich connections within main components of the Earth's system.

narrow window of the planet's history. The Earth is continuously experiencing natural events such as geological and tectonic processes, climate change, and biological and chemical activities. Although the instruments to record such events were not available before the 17th century, various natural and complex formations, such as stalagmites, marine and lake sediments, and trees, have recorded such events in their structures as proxy records. Investigating these archives to reveal the hidden preceding events helps us understand the dynamics and predict the oncoming behavior of the associated natural events on the Earth. For this purpose, paleo-climatology, a field of climate science that works to understand ancient climates without direct measurement, has reached sufficient maturity to reveal significant climate periods, such as glaciations and abrupt global temperature rises, by dating and analyzing the proxies [6].

Whereas paleoclimate variations as derived from the geoscientific archives are only estimates and contain a degree of uncertainty, the significant climate periods of the driving processes such as the Milankovich cycles can be determined with high accuracy because the equations of motion for the dynamics of the Earth's orbit in space can be solved with a reasonable approximation using Hamiltonian mechanics. However, the celestial sign of objects in the solar system is, in general, a many-body system in which the planets' gravitational fields mutually influence their orbits around the Sun. Solving such a many-body problem (and even that of a three-body system) is not simple and was at the forefront of science for a long time [7]. In this spirit and in honor of the 60th birthday of the King of Sweden, Oscar II, in 1887, a prize for solving the many-body problem was announced. The French mathematician Henry Poincaré finally won this prize with his seminal work on the three-body system and discovering the chaotic nature of the planets' orbits [8]. In this work, he proved an important theorem that affects the recurring orbits of the interacting objects in a celestial system and is also a fundamental property of many complex dynamical systems: the now well-known recurrence theorem, which states that a (conservative) system recurs infinitely many times as closely as one wishes to its initial state. The property of recurrence is not only of fundamental importance in the study of dynamic systems; it is also a fundamental principle in Earth sciences at all temporal and spatial scales.

6.2.2 Nonlinear concepts

The vigorous progress in exploring nonlinear dynamics in the 1980s and 1990s opened new doors for a more appropriate analysis of complex nonlinear systems, such as lasers, the human brain, power grids, and the Earth's system [9]. Techniques for estimating fundamental characteristics of nonlinear systems, such as fractal dimension, Lyapunov exponents, Kolmogorov entropy, and Hurst exponents, were developed and applied to various disciplines [10]. However, these methods are mainly helpful in low-dimensional processes and are not appropriate for understanding the Earth's system from data.

Other essential concepts, such as recurrence plots [11, 12], synchronization [13], wavelets [14], and complex networks [15], have been developed to explore dynamical and structural properties in high-dimensional spatiotemporal systems. They have been proven to be very promising even for the study of the Earth's system. In the following subsections, we describe basic nonlinear concepts and present some paradigmatic applications in Earth sciences.

6.2.2.1 Multiscaling

Various Earth processes are assumed to be *scale-invariant* [4]. An essential law is the size distribution of natural events, meaning that prominent events are less frequent when compared to smaller ones. Deriving an adequate size distribution of natural events would take into account the rarity and likelihood of a specific event. Hence, one major challenge of studying the occurrence, frequency, and intensity of climate-driven natural extremes and natural hazards is determining these events' spatial and temporal scaling to derive adequate risk estimates. One way to analyze the scaling of natural hazards is to use the frequency–size distribution $p(x)$ (x stands, e.g., for landslide area). For instance, $p(x)$ of landslides follows a power law probability density function in an area with arbitrary dimensions independently of their source mechanism (e.g., earthquake- or rainfall-induced):

$$p(x) = (\alpha - 1)x_{\min}^{\alpha-1}x^{-\alpha} \tag{6.1}$$

with α the power exponent, valid for $x \geqslant x_{\min}$ [16].

Similarly, the famous Gutenberg–Richter power law [17, 18] scales the seismic activity to assess earthquake hazards for different events magnitudes m. It states that earthquake magnitudes m are distributed exponentially as

$$\log N_{m\geqslant M} = a - bM \tag{6.2}$$

where $N_{m\geqslant M}$ is the number of earthquakes with magnitude $m \geqslant M$, a is a constant, and b is the scaling parameter. The scaling parameter b determines the relative frequency of small and large earthquakes. The estimation of b is around 1.0, with deviations up to 30% in seismically active regions [19]. A real example of this particular case is presented in section 6.2.3.1.

Information about the Earth's system's processes can be stored and perceived differently at multiple scales. The information observed at one scale often cannot be directly used as information at another. Scaling approaches address the changes at the measurement scale and play an essential role in Earth sciences by providing information at the scale of interest.

Determining scaling properties of geophysical variables provides an alternating way to obtain information about the associated process. The processes with similar statistical properties at different scales are said to be self-similar, which can be described mathematically as [20]

$$\phi(x) = \lambda^{-\beta}\phi(\lambda x) \tag{6.3}$$

where x is the finer spatial resolution (scale), β is the scaling exponent, λ is the ratio of the large resolution, λx to the small resolution x, and ϕ is the geophysical property or variable of interest. A field is said to be spatially scaling with respect to the moment, q, if the following relationship holds [21]:

$$E\left[\left(\phi_\lambda\right)^q\right] \propto \lambda^{K(q)} E\left[(\phi_1)^q\right] \tag{6.4}$$

where $K(q)$ is the scaling exponent associated with the moment of order q. If the exponent $K(q)$ is linear with regard to q, the process has simple scaling. On the other hand, if the scaling exponents, or slopes, are a nonlinear function of q, then the process is said to be *multiscaling*. This concept of scaling and multiscaling has been used widely in many scientific fields, including hydrology and ecology. For instance, wavelet analysis can decompose high-resolution nonstationary spatial information into nonstationary fields of increasingly coarse spatial scales [22]. The wavelet and the corresponding scaling function are a function to decompose spatial information into directional components explained by the wavelet coefficients.

6.2.2.2 Recurrence analysis

The seminal work of Poincaré in 1890 [8] played a central role in the qualitative theory for nonlinear dynamics (see section 2.1). Poincaré presented a method that provides a local and global analysis of nonlinear dynamical systems by the Poincaré recurrence theorem and stability theory for fixed points and periodic orbits. This theoretical finding is compellingly confirmed by the real world, where recurrences can be observed in our daily life and across all scientific disciplines. Therefore, the investigation of recurrences has attracted much attention, and several approaches have been developed for this purpose.

Among the various methods for studying recurring processes, power spectrum analysis is one of the best known and most widely used techniques for identifying periodicities in time series [23]. Wavelet analysis reveals similar information, additionally providing the change of the detected periods over time (see section 2.1). Coming from the theory of dynamical systems and based on Poincaré's recurrence theorem, the *recurrence plot* (RP) is another fundamental approach that can be used to investigate recurring features in time series and even in spatial data [11, 12]. In a given m-dimensional phase space, two neighboring points are called recurrent if the distance between their state vectors is closer than the threshold ε. Formally, for a given trajectory $\boldsymbol{x}_i$ ($i = 1,\ldots, N$, $\boldsymbol{x} \in \mathbb{R}^m$), the recurrence matrix $\boldsymbol{R}$ is defined as

$$R_{i,j} = \begin{cases} 1, \text{ if } \left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\| \leqslant \varepsilon \\ 0, \text{otherwise} \end{cases} \tag{6.5}$$

where $\|\cdot\|$ is a norm of the adopted phase space. The graphical representation of the recurrence matrix $\boldsymbol{R}$ is the RP (figure 6.2). RP of different dynamical behavior represents different particular features (figure 6.2). Such differences can be quantified with the measures of recurrence quantification analysis, such as determinism (the fraction of recurrence points that form diagonal lines in the recurrence plot), laminarity (the fraction of recurrence points that form vertical lines), and recurrence rate (the percentage of recurrence points in a recurrence plot). These measures are used to find changes in the dynamics of a process (e.g., in climate), to classify the dynamics (e.g., random, chaotic, regular) [12], or to identify interrelationships and coupling directions in coupled systems [24].

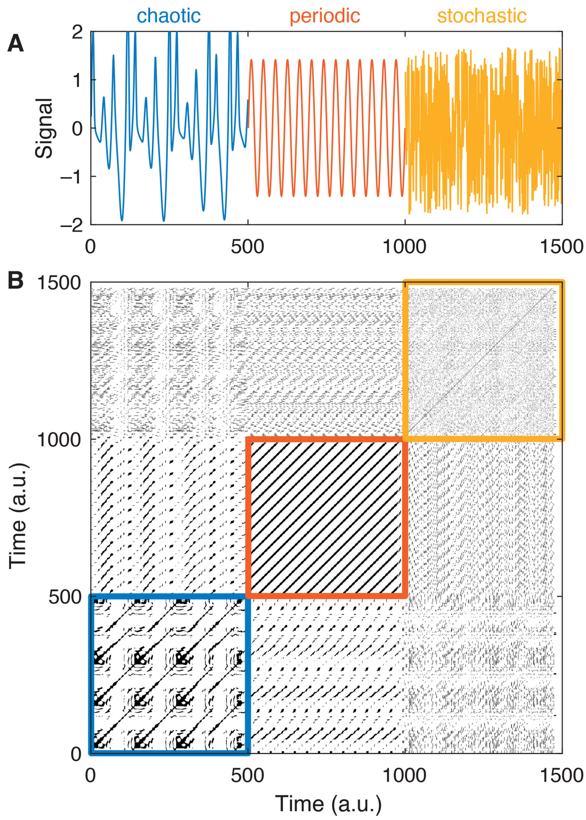


Figure 6.2. (A) Time series representing switching between different dynamical regimes, from chaotic via periodic to stochastic, each lasting 500 time steps. (B) A recurrence plot (RP) represents the recurrence of a state at a given point in time (x-axis) at another point in time (y-axis). Different dynamics cause typical recurrence patterns, which can be used to detect these changing dynamical behaviors. Continuous long diagonal lines in the RP indicate the periodic window, shorter diagonals show the chaos, and single points appear in the stochastic part.

6.2.2.3 Complex networks and event synchronization

Essential challenges in climatology are quantifying the spatial extent of climate extremes and early forecasting procedures of their dynamical behavior. Such forecasting relies predominantly on numerical models which solve physics-based coupled systems of partial differential equations. Starting with Richardson in the 1920s, it has been a long way to the first successful prediction in 1950 and eventually to today's highly sophisticated general circulation and Earth system models. Despite multiple efforts using these methods, their predictive power, especially for extreme events, can be rather limited. A primary reason for this is that in particular long-range interactions, called *teleconnections*, and their interaction with more regional interactions may not be well represented or may even be absent in such models.

Therefore, a quite different approach has been suggested: a network-based presentation of climate phenomena called *climate networks*. The main idea is to get additional information by capturing the evolving interactions of different locations, regarded as nodes, through similarity measures, such as the Pearson correlation, mutual information, or the Granger causality, from spatiotemporal observational data. An important description of such similarity of strong events is the *event synchronization* approach [25], inspired by Christiaan Huygens' detection of synchronization in the 17th century. Here, we consider the occurrence of extreme events, such as rainfall, in a synchronized manner at different locations, even faraway ones.

The final complex network is then represented by an adjacency matrix $\boldsymbol{A}$, which encodes the links between the nodes i and j as follows:

$$A_{i,j} = \begin{cases} \text{nonzero}, & \text{if variability at node } j \text{ is similar (or synchronized) to node } i \\ 0, & \text{otherwise} \end{cases} \tag{6.6}$$

The value of the elements of $\boldsymbol{A}$ represents the weight of the link obtained from quantifying similarity (figure 6.3).

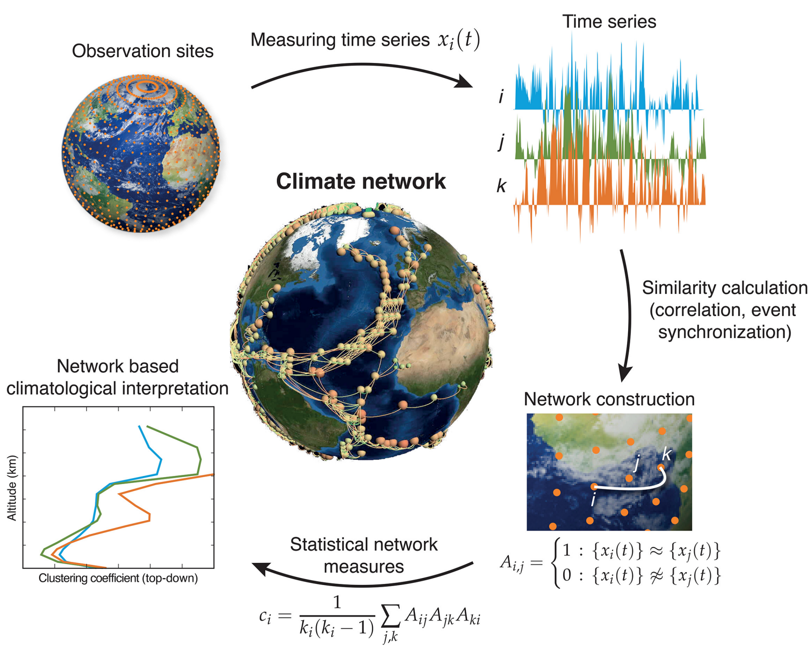


Figure 6.3. The climate network framework as a tool for prediction. Observational data of physical quantities, such as temperatures, are available at different geographical locations. These data can be used directly or via a reanalysis (numerical weather model) which assimilates and maps them onto a regular grid. Thus, a time series of the regarded physical quantity is available for each climate network node (observational site or reanalysis grid point). Cooperativity between nodes can be detected from the similarity in the evolution of these time series and translated into links connecting the corresponding nodes. The links or their strengths may change with time. These nodes and their links constitute the evolving climate network represented by the adjacency (connectivity) matrix $\boldsymbol{A}$ (equation (6.6)). The analysis of this network can enable early predictions of climate phenomena and provide insights into the physical processes of the Earth's system.

There are various generalizations of this construction, particularly to emphasize multilayer networks, which enable variables from different subsystems.

The reconstructed adjacency matrix $\boldsymbol{A}$ allows us to calculate standard network measures such as degrees, clustering coefficients, or betweenness and to identify teleconnections. It has been shown recently that climate networks provide ideal tools for exploring even large amounts of climate data to uncover spatiotemporal patterns, leading to new physical insights into the climate system [1]. Moreover, they have a strong predictive potential; that is, they enable the development of new forecasting methods. Examples of up-and-coming applications are given in sections 3.3–3.5.

6.2.3 Applications of nonlinear dynamics in the Earth's system

Vigorous progress in nonlinear science contributed to detecting, attributing, and understanding the Earth's system, reducing uncertainties, and projecting future climate changes. In this section, we discuss some significant contributions of nonlinear physics in Earth system sciences.

6.2.3.1 Earthquakes and the Gutenberg–Richter law

A proper fitting of the power law is essential to study most natural hazards, particularly earthquakes (equation (6.1)). The Gutenberg–Richter law (equation (6.2)) represents scaling in earthquakes, as power law distribution makes it scale-invariant. An example of the scale parameter b for central California for 20 years (2001–20) is illustrated in figure 6.4. The California region has a b of 1.0, which is as per the global average, meaning that central California has the same relative

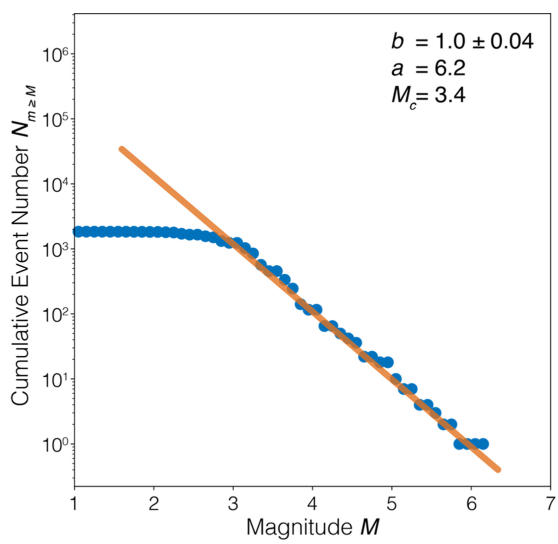


Figure 6.4. Frequency–magnitude distribution for earthquakes in central California between 2001 and 2020. The red line shows a fit to the cumulative frequency and has a slope (b-value) of 1.0. The magnitude cutoff, $\boldsymbol{M_c = 3.4}$, is used for estimating the scaling parameter b.

frequency of small and large earthquakes. However, the magnitude threshold parameter M_c must be selectively applied above crossover magnitude for larger earthquakes with significant seismic moments [26]. The Gutenberg–Richter law accurately describes the shallow seismicity. However, it is not the only scaling law for all levels of earthquake events; the distribution of deeper earthquakes was observed to follow a bimodal (multiscaling) pattern [27].

It is also crucial to accurately estimate scaling parameter b (equation (6.2)) from the earthquake events to characterize the seismicity activity (see section 2.1) sensitively. There is an inverse correlation between b and the differential stress, which was revolutionary in that b can act as an indicator of stress accumulated around the fault volume [28]. This observation was used in the study done before and after the vast 2011 Tohoku-Oki earthquake with a high slip area, where an increase in b is observed as a large amount of stress was released [29]. Another use for this observation is studying the structural anomalies in the crust and identifying the volumes of magma in an active volcano. A study performed at two active volcanoes [30], Mt. St. Helens and Mt. Spurr, shows a relatively high b ($\geqslant$1.3) due to the presence of material heterogeneity and high thermal gradient. This high b is why these volcanoes are less likely to host large earthquakes but frequent small ones. A typical intraplate b is around 0.8, making intraplate regions prone to large earthquakes over a short recurrence time. However, the scaling parameter b is not the perfect parameter to measure seismicity at all magnitude scales. The tail of the $\log(N_m > M)$ versus M relation holds for only a certain range of magnitudes. A nonlinear fit is a better approximation for smaller ($M_c \leqslant 3.4$) and larger ($M_f \gtrsim 7$) magnitudes. A reason for the deviation from the power law for earthquakes smaller than $M_c \leqslant 3.4$ (figure 6.4) is the incompleteness of catalogs. For large earthquakes, a reason is the saturation of the magnitude scale and the long recurrence time; they are missing from the catalogs because they are often too short.

A high scaling parameter ***b*** indicates a lower chance of observing significant seismicity while the frequency of small earthquakes is high. However, smaller-magnitude events are observed much less often than indicated by b due to insufficient seismic network coverage.

6.2.3.2 Recurrence plot application

Recurrence is a fundamental principle in Earth sciences at all temporal and spatial scales, from the key principle of the doctrine of uniformity, over the rock cycle, glaciation cycles, and active geysers, to alternating sediment layers, to mention only a few. One crucial phenomenon with complex recurrence patterns is climate. One of the primary drivers of climate is solar insolation, modulated by mutual variations of the Earth's orbit around the Sun and the tilt of the Earth's axis, which are responsible for seasons, changes in global temperature, and glaciations. This influence was discovered in the first half of the 20th century by investigating annually layered lake sediments [31] and considering the Earth's orbital parameters [32].

Recurrence plots (see section 2.2) provide a powerful framework to study the dynamics of the climate by their recurrence properties. As an application, the dynamics of the Cenozoic climate will be investigated by recurrences properties in a selected paleoclimate proxy record. Such studies are essential to advance our understanding of the past and will help to improve climate models to better forecast future climate change and its impacts, as well as increasing our understanding of climate dynamics.

Calcareous lake sediments, speleothems, and benthic foraminifera store environmental conditions by changing their geochemical and petrographic composition. The study of stable isotopes is an active field to derive past environmental and climatic conditions. For example, the temperature-dependent fractionation of oxygen isotopes is the key to reconstructing global seawater temperatures and ocean circulation by using planktonic and benthic foraminifera. Ongoing deep ocean drilling programs and novel quantitative methods such as clumped isotope thermometry provide new insights with improved quantification, increasing temporal resolution, and ever-smaller time uncertainties. The recently developed temperature reference curve for the Cenozoic [33] is an example with a temporal resolution of up to 2000 years and covering 66 million years. This period is crucial because it provides an analog of future greenhouse climate and how (and which) regime shifts in large-scale atmospheric and ocean circulation can be expected in a warming world [34]. The outstanding high resolution of this record allows study and comparison of recurrence properties of selected time intervals. The recurrence plot indicates the different climate regimes of hothouse, warmhouse, coolhouse, and icehouse by their very distinct recurrence pattern (figure 6.5). During the Miocene (18–14 Ma ago), the climate was in a warmer state more similar to the warmhouse than the coolhouse, visible by some recurrences linking this period to the late Eocene. The fine-scale pattern of the recurrence plot reveals more details, such as the change from the 41-ka cycles to 100-ka Milankovich cycles of glaciation during the mid-Pleistocene transition.

Recurrence analysis of climate time series indicates different dynamical regimes, such as chaotic or predictable dynamics, thus enabling detection of critical transitions between different climate periods.

6.2.3.3 Extreme rainfall teleconnections and monsoon prediction

The Indian summer monsoon is an intense rainy season lasting from June to October. The monsoon delivers more than 70% of the country's annual rainfall, which is India's primary source of freshwater. Although the rainy season happens every year, the monsoon onset and withdrawal dates vary within a month from year to year. Such variability strongly affects the life and property of more than a billion people in India, especially those living in rural areas and working in the agricultural sector, which employs 70% of the entire population. So far, only Kerala in South India receives an official monsoon forecast two weeks in advance, while the other 28 states rely on the operational weather forecast of about five days [35]. A much better forecast has been recently reached by combining two nonlinear concepts: complex climate networks and a tipping element approach.

In the first step, from rainfall data from the Asian Precipitation Highly Resolved Observational Data Integration Towards the Evaluation of Water Resources

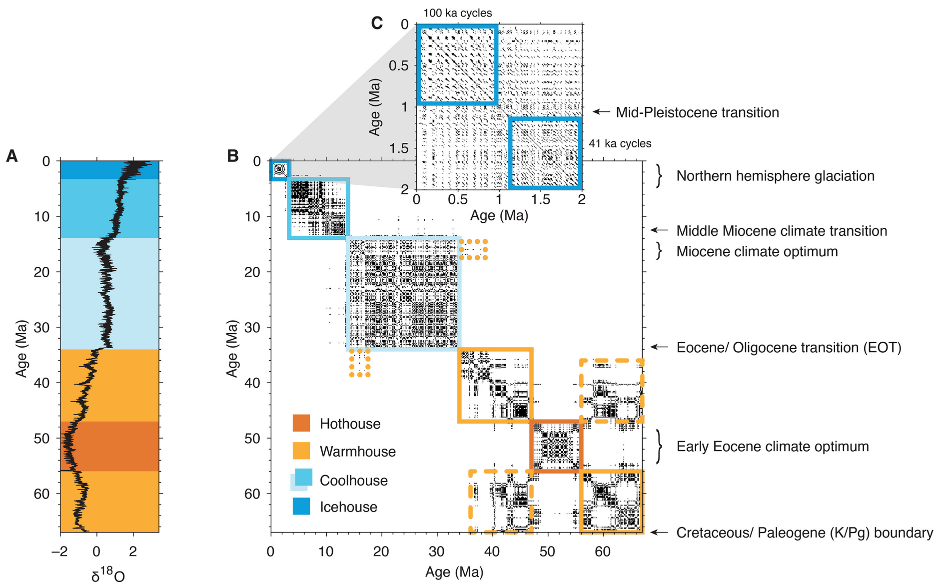


Figure 6.5. RP of a paleoclimate time series. (A) Paleoclimate variation indicated by oxygen isotope measurements from marine sediments (CENOGRID). Lower values correspond to a warmer global climate. (B) The RP indicates the different climate regimes of hothouse, warmhouse, coolhouse, and icehouse by their very distinct recurrence pattern. During the Miocene (18–14 Ma ago), the climate was in a warmer state more similar to the warmhouse than the coolhouse, visible by some recurrences linking this period to the late Eocene (marked by the dashed box). (C) The fine-scale pattern of the RP reveals more details, such as the change from the 41-ka cycles to 100-ka cycles of glaciation during the mid-Pleistocene transition.

(APHRODITE) and the high-resolution satellite product Tropical Rainfall Measurement Mission (TRMM) 3B42 dataset, complex networks were retrieved via the event synchronization technique (see section 2.3). This exploratory network-based analysis of extreme rainfall across the Indian subcontinent enabled for the first time the identification of critical geographical domains displaying far-reaching links, influencing distant grid points [36]. In particular, North Pakistan and the Eastern Ghats turn out to be crucial for the transport of precipitation across the subcontinent.

In the second step, a tipping elements approach of the measured daily mean air temperature and the relative humidity at these two sensitive regions allowed us to uncover the critical nature of the spatiotemporal transition to the monsoon. It was especially found that the temporal evolution of the daily mean air temperature and the relative humidity exhibits critical thresholds on the eve of the monsoon. A highly developed instability occurring in these regions creates the conditions necessary for spatially organized and temporally sustained monsoon rainfall.

Based on this knowledge, a scheme was developed for forecasting the upcoming monsoon onset in the central part of India 40 days in advance, thus considerably improving the time horizon of conventional forecasts. The new scheme not only has proven its worth (73% of onset predictions have been correct) in retrospect (for the years 1951–2015) but also has already been shown to be successful in predicting future monsoons five years in a row since its introduction in 2016. The methodology appears to be robust under climate change and has proven its skill also under the extreme conditions of 2016, 2018, and 2019.

Further successful applications of this network-based concept are El Niño forecasts beyond the spring barrier, predicting droughts in the central Amazon 12–18 months in advance, and forecasting extreme rainfall in the Eastern Central Andes [37].

Thus, a network-based analysis of climate data can provide predictive power for mitigating the global warming crisis and societal challenges.

6.2.3.4 Understanding landslide distributions

As explained in section 6.2.2.1, successfully fitting a global power law distribution (equation (6.1)) to landslides would help us to understand whether we lack information in hazard and risk models. However, the distribution of spatial landslides follows a power law distribution. Just as in the case of the Gutenberg–Richter law, the power exponent is valid to a minimum value (equation (6.1)) [16], and the rollover below the minimum is found in two different forms: (1) the double Pareto distribution and (2) the inverse Gamma distribution according to different studies [38, 39]. Like other universal scaling laws [40], it is expected to have a universal power exponent for the landslide events. However, a lack of data makes studying the problem impossible at a better resolution, especially at the function's tail [38, 39]. Most studies rely primarily on landslide inventories collected after a significant landslide triggering event, such as the 1994 Northridge earthquake (M_W 6.4). Landslides have also been found to exhibit temporal scaling or clustering besides spatial and geometric ones. Although some studies suggest a global power exponent $\alpha = 2.3 \pm 0.6$, the physical process is not known to implement a functional probabilistic multihazard assessment [41].

Besides the power law–based approximation models, ample practice has offered linear solutions to study natural hazards, making a nonlinear application redundant. An example is Newmark's sliding block analysis. It estimates the displacement potential of hillslopes under seismic loading (i.e., acceleration). This hypothetical displacement aims to indicate the likelihood of failure under seismic loading as a function of hillslope inclination and seismic acceleration. For example, landslides related to the 2016 Kumamoto earthquake (M_W 7.1) caused significant damage, especially to infrastructure such as highways (figure 6.6(A)). Although landslide locations correlate well with the seismic waveforms based on a physics-based ground motion model [42], the Newmark's distances highlight particularly elevated gradients in the landscape (figure 6.6(B)).

Rainfall decreases the slope stability by altering cohesion, elevating the landslide susceptibility in most cases. In some other cases, rainfall could also mobilize the superficial surface material leading to the debris flows. However, in contrast to an earthquake, rainfall is not introducing a direct force on the hillslopes to estimate rainfall impact on landslides. Hence, most of the time, statistical methods are applied to forecast rainfall-induced landslides. One standard tool is to use statistically derived rainfall intensity–duration thresholds above which landslides are

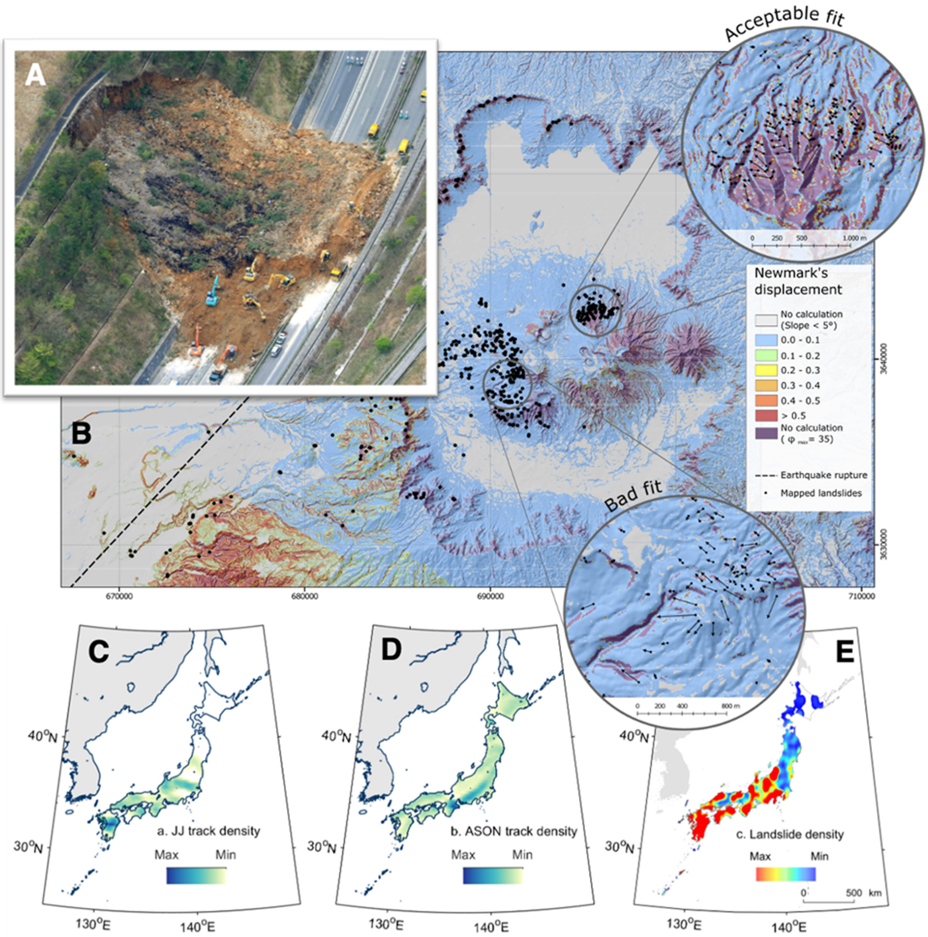


Figure 6.6. (A) Example of a cut slope failure by the Oita Expressway following the 2016 Kumamoto earthquake (M_W 7.1). The photo is taken from Dave Petley's landslide blog (https://blogs.agu.org/landslideblog/2016/04/18/kumamoto-Earthquake-1/). (B) Newmark's displacement of the 2016 Kumamoto earthquake (M_W 7.1) in Kyushu, Japan (UTM-52). In certain regions, the elevated displacement correlates well with the mapped landslides, while in some others, it is relatively poor. The concentration of extreme precipitation streamlines during (C) June and July (JJ), and (D) August to November (ASON), normalized by cumulative above 95% extreme rainfall for the same period between 1998 and 2015 based on TRMM (Tropical Rainfall Measurement Mission) rainfall estimates. (E) Normalized rainfall-triggered spatial landslide density -weighted by log-transformed landslide volumes calculated from an inventory of 4744 events and smoothed by kernel density estimation onto a 5 × 5 km grid by [44]; white areas have no data.

triggered. The logic behind is that high-intensity rainfall triggers landslides and moderate intensity, but long-duration events would increase the landslide susceptibility. Therefore, several spatial classification models are developed to try to relate landslide activity to rainfall distribution.

Another notorious example is that the extreme rainfall flux over a region during tropical storms, as previously explained, might already highlight the landslide-prone regions on large spatial scales. It is possible to estimate or cluster the rainfall motion over large areas, such as countries or continents, by blending event synchronization and complex network methods (see section 2.3). These results can help track landslide activity along the path of extreme rainfall. As an application, the extreme rainfall trajectories over the Japanese archipelago were estimated using event-synchronization [43]. The density of extreme rainfall tracks aligns well with the landslide distribution (figures 6.6(C)–(E)).

The power law distribution can model landslide distributions, and by using nonlinear methods such as event-synchronization, it is possible to describe spatial landslide distributions as a function of rainfall distributions.

6.2.3.5 Multiscale sea surface temperature (SST)

Climatic systems are complex systems composed of multiple feedback loops and interactions. In such systems, the coupling between climate variables takes place at different time and spatial scales. Untangling this multiscale variability and interactions of a climatic process is vital, as it would improve the understanding of global climate and its variability. Hence, climate networks are constructed (section 2.3) at different time scales considering each sea surface temperature (SST) grid cell as a node, and edges are created between all pairs of nodes based on statistical relationships. First, SST data are decomposed at different time scales using wavelet (section 2.1), and then the Pearson correlation between all pairs of nodes is calculated at a corresponding time scale. Finally, significance-based pruning is applied to retain only highly correlated edges in the network. The network is constructed by applying a 5% link density threshold, which is well accepted for the network construction. Multiple testing was employed to avoid false links.

The network visualization of the original SST data (all scales) reveals short-range and long-range connections between various regions of the Earth. As at a finer scale, there is no significant correlation, and that is expected, since we have removed the annual cycle using anomalies. Interestingly, at 8–16 months, we observe mainly two zones with many significant correlations in the equatorial Pacific and Indian Ocean dipole, which are known to affect each other via the atmosphere (figure 6.7(A)). In the next period of

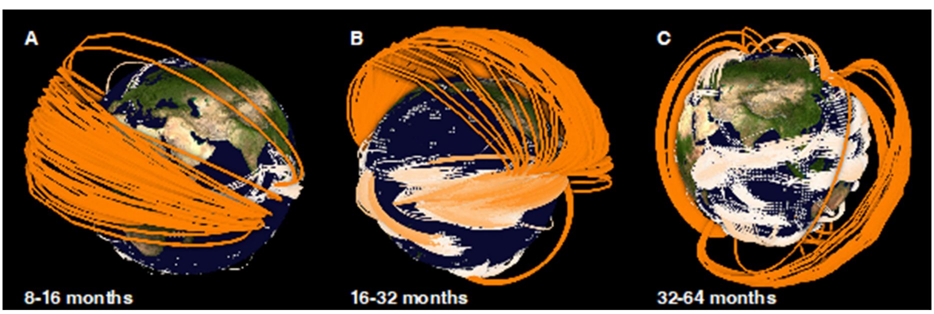


Figure 6.7. Spherical three-dimensional globe representation of the long-range teleconnections at different timescales in a sea surface temperature network. Reproduced from [45] CC BY 4.0. Edge color represents the geographical lengths.

32–64 months, these patterns become more prominent as known ENSO events act on scale up to two years (figure 6.7(B)). There is a link between SST in the Southern Ocean to ENSO events via the Southern annual mode, that is, the north–south movement of the westerly wind belt that circles Antarctica (figure 6.7(B)). The three-dimensional visualization (figure 6.7(C)) shows several links from the North Atlantic to the South Atlantic. This negative correlation likely exhibits the seesaw response due to the transport of heat from the Southern Ocean to the North Atlantic via the Atlantic meridional overturning circulation (AMOC) [46]. If the AMOC is stronger than it was before 2000 (as it has been in the period since the year 2000 compared to the years before [47]), more heat is transported towards the North, which leads to a cooling in the Southern Ocean and warming in the subpolar North Atlantic.

Multiscale analysis of climatic processes helps to uncover the time scales of interaction and feedback in the climate system that may be missed when processes are analyzed at one timescale only.

6.2.4 Outlook

We have shown that basic concepts of nonlinear physics and complex systems science have a strong potential for treating important problems in Earth systems sciences. We have argued that they complement established concepts with new possibilities to reveal entire causal chains of complex phenomena in the Earth's system, primarily to reveal new precursor processes of extreme events.

However, it is essential to emphasize that these interdisciplinary approaches are in their infancy and the subject of ongoing research. There are various open challenges in the realm of methodological nature and applications. Some of them are as follows:

- There is a growing recognition in the scientific community and more broadly that the Earth's functions have to be regarded as an interconnected complex system with properties and behaviors characteristics of the system as a whole. These include tipping points, critical thresholds, 'switch' or 'control' points, strong nonlinearities, teleconnections, chaotic elements, and uncertainties of different origins. Understanding the components of the Earth's system is important; however, that is insufficient for understanding the functioning of the Earth's system as a whole. Humans are now a significant force in the Earth's system, altering key process rates and absorbing global environmental changes. Human activities' environmental significance is so profound that the current geological era is called the Anthropocene [48]. Therefore, there is a strong need to develop a complex global model involving Earth system dynamics, human activities, and environmental boundaries to systematically study the planetary boundaries and tipping points and uncover fundamental principles.
- An important task is to improve our capabilities regarding data-driven inference of governing principles to reach a deeper understanding of the connection between the microscopic dynamics of the constituents of the Earth's system and their nonlinear interactions on the one hand and the dynamics emerging from these interactions at the macroscopic level on the other hand.

- Combining traditional physics-based modeling and statistical approaches with state-of-the-art machine learning (ML) techniques is necessary to efficiently include the huge amount of available data in a model. However, we would like to emphasize that neither an ML-only nor a scientific knowledge-only approach is sufficient for complex Earth system applications. Hence, we must explore the continuum between mechanistic and ML models, where both scientific knowledge and data are integrated synergistically [49, 50]. This approach has picked up momentum just in the last few years [50] and is being pursued in Earth system science [51], climate science [52], and hydrology [53].
- A key driver of further advances is the desire to improve predictions of the behavior of complex systems and especially—for example, in the context of the ongoing global warming driven by the anthropogenic release of greenhouse gases—of the response of complex systems to time-varying external forcing.
- The study of surface processes with nonlinear tools is still not common. Combining nonlinear approaches with linear methods could advance the existing forecasting schemes, especially predicting extreme events. The European floods in summer 2021, which claimed 184 lives in Germany alone [54], are a matchless example that emphasizes that more effort has to be placed to forecast extreme incidents to prevent life loss.
- Climate-driven hazards are rarely the output of a single system. Many of those are in the form of hazard cascades. For example, extreme rainfall initiates a flash flood, high waters lead to carving the river banks and trigger landslides, and dislocated loose landslide mass mixes with the high waters and is transported downstream as a debris flow. However, most of the research that links urban interaction with climate-driven natural hazards consists of empirical studies. Only recently, the first numerical model (CHASM) has been able to describe the informal housing-related changes in the topography and link it to the occurrence rates of landslides [55]. Models such as CHASM could connect different earth systems. Blending such physics-based models and a nonlinear causation metric such as event synchronization as a preceding step could enhance our capacity to forecast extreme rainfall-driven natural hazard cascades, such as flash floods and landslides.
- The simultaneous occurrence of two or more natural extremes affects society much more strongly than their univariate counterparts do [56]. For instance, a hazard resulting from a summer that is both dry and hot is higher than that resulting from a univariate drought extreme, given that the hot, dry summer has a severe impact, such as a reduction in agricultural productivity, irretrievable loss to property and health, and damage to natural ecosystems and public infrastructure. These manifold extremes are called compound extremes or compound events [57]. The investigation of compound extremes has received less attention so far; nevertheless, it has recently gained significant momentum across the globe [56, 58–60].
- Overall transient central components of Earth systems, such as temperature and rainfall, and their effects on other processes, such as concomitant natural hazards of droughts or landslides, should be emphasized and studied using more recent comprehensive data.

6.3 Physics fields with relevance for energy technologies

6.3.1 Solar energy

Robert Pitz-Paal[1] and Bernd Rech[2]
[1]German Aerospace Center (DLR), Cologne, Germany
[2]Institut für Silizium-Photovoltaik, Berlin, Germany

6.3.1.1 The solar resource and the global potential of solar energy technologies

Solar energy is the most abundant resource of energy in the world. With approximately 23 000 TW per year reaching the world's surface, solar energy exceeds the current world energy use by approximately a factor of 3800. However, its average energy density is low (≈160 W m^{-2}) compared to that of other energy sources. Less than 1% of solar energy is converted to other renewable energy sources such as wind, biomass, or hydro power. Fossil and nuclear reserves together account for less than 10% of the yearly solar resource. Based on these numbers it is not surprising that sustainable energy scenarios consider solar energy to become one of or even the major energy source [61].

However, two major challenges need to be overcome: First, the low energy density requires a cost-efficient technology for the collection. The second challenge is the inhomogeneous distribution of solar radiation in space and time (see figure 6.8), which requires cost-efficient energy transport and storage technology.

Two different technical concepts have been developed successfully over the last five decades that will be discussed in this section.

Photovoltaics (details in box 6.1) converts high-energy photons into an electric current, taking advantage of Einstein's photoelectric effect in a semiconductor. Cost efficiency is achieved by mass-produced semiconductor devices that are electrically connected to collect energy over large surface areas. Forms of electrical energy storage such as batteries are required to provide electricity on rather short-term

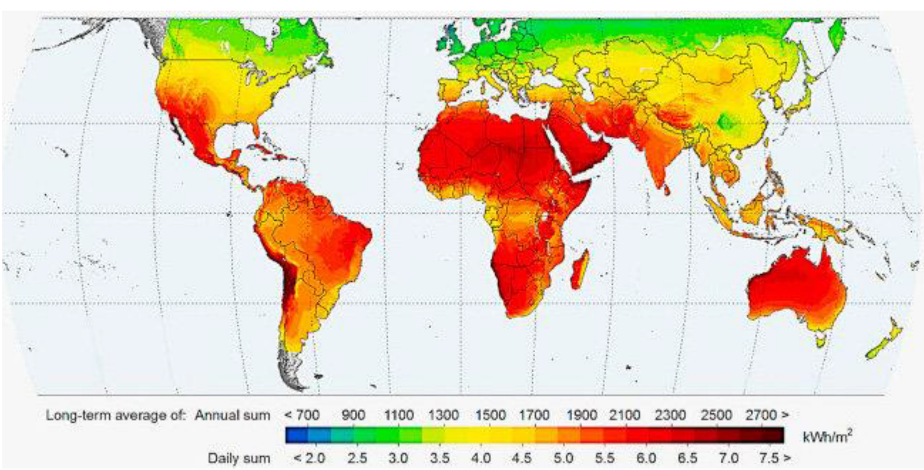


Figure 6.8. Global map of annual solar horizontal radiation [62]

demand, while the production of hydrogen via electrolysis opens the path towards long-term storage.

Box 6.1 Photovoltaics.

Solar cells directly convert light into electricity. The thermodynamic efficiency of the conversion process is determined by the surface temperature of our Sun with approximately 5800 K enabling an efficiency potential above 80%. The basic working principle of solar cells relies on a semiconductor material which is embedded between selective contacts, collecting the electron and holes which are generated in the semiconductor by the photoelectric effect (also see figure 6.9). The prerequisite is

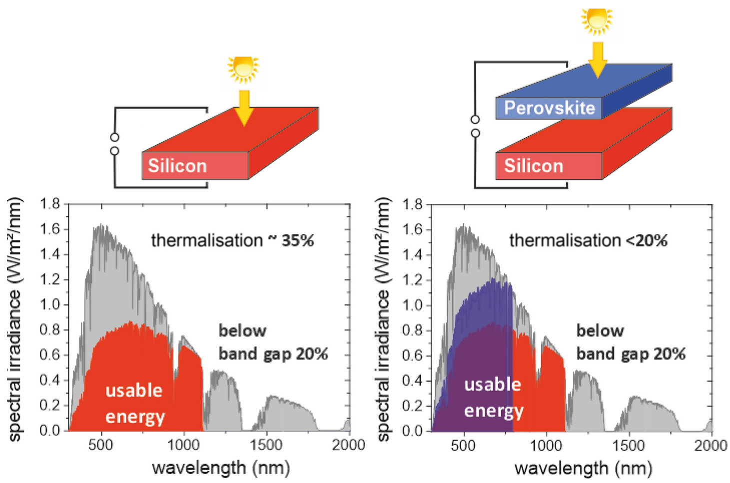


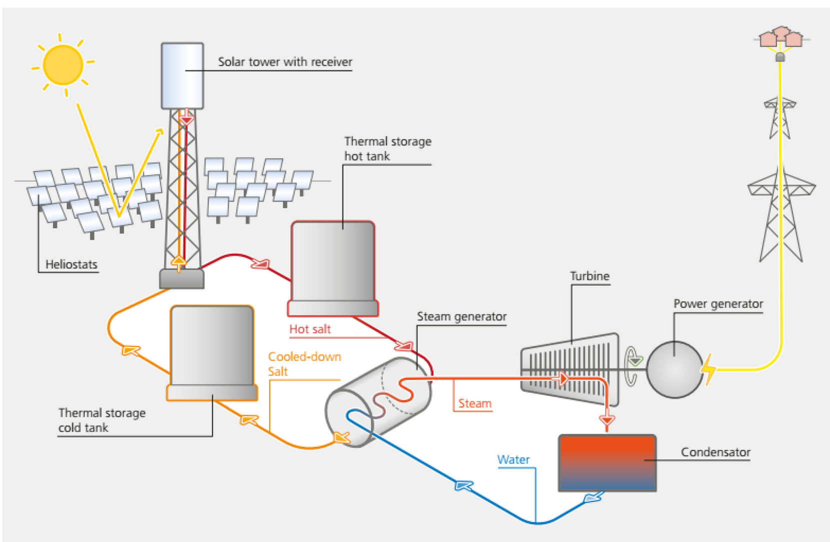
Figure 6.9. Working principle and usable energy of a single junction and a tandem junction solar cell. The solar irradiance with an energy higher than the bandgap energy E_g of the semiconductor is absorbed and converted into electrical energy. The photon energy higher than E_g is lost through thermalisation. The multijunction solar cell architecture (in this case two junctions combining silicon and a higher band gap perovskite base semiconductor) utilises the higher photon energy range (lower wavelength regions) at higher voltages, thus enhancing the overall solar conversion efficiency.

that the photon energy is larger than the band gap of the semiconductor. Note that the photo-excited electrons and holes lose their excess energy (photon energy minus band gap energy) in a very fast process called thermalization. Further loss processes are reflection losses at the front side of the solar cells and recombination losses when carriers recombine before they reach the carrier-selective contacts. Shockley and Queisser [63] derived an efficiency limit for a single junction solar cell of a few percent above 30% under one sun and 40% under highly concentrated sunlight. The theoretical limit for multijunction cells (shown as a tandem cell combining two semiconductor materials) with an infinite (or very large) number of component cells, each adapted to a specific part of the solar spectrum, is higher than 80% because virtually no excess photon energy is lost. Today, the world record cell efficiency measured under concentrated sunlight with a multijunction solar cell is 47.1%, while the benchmark for crystalline Si cells—the PV mainstream today—is 26.7% in conversion efficiency for laboratory-type small-area cells [64].

Solar thermal technology (details in box 6.2) is based on the broadband absorption of solar photons that are converted to heat. In combination with concentrating devices a high-temperature heat source can be established that can power a thermodynamic cycle to generate mechanical energy used for electricity production. In this concept, heat is stored rather than electricity to provide electricity on demand, as this can be realized at significantly lower cost. High-temperature thermal systems rely on low-cost sun-tracking mirrors spread over large surfaces that concentrate the energy on a small receiver device, where it is absorbed and converted to heat. Concentration devices can use only the direct component of the solar radiation, as diffuse radiation (i.e., photons that were scattered on the way through the Earth's atmosphere) cannot be concentrated, according to the laws of physics. This limits the cost-efficient application of this technology to sites with a high direct component, such as we find in the Earth's sunbelt.

Box 6.2 Solar thermal power.

Solar thermal power plants use mirrors to gather direct sunlight and convert it into heat. This is used to generate steam to operate a turbine, which in turn drives a generator to convert kinetic energy into electrical energy. The integrated heat storage system enables the power plant to accurately generate electricity when needed without being affected by fluctuations in solar radiation intensity throughout the day (see also figure 6.10). Additionally, the use of fossil or renewable fuels can compensate for



How a solar thermal power plant works, shown here for a solar tower power plant. Image: DLR

Figure 6.10. Schematic of a solar thermal power plant: a concentrator composed of heliostats, a receiver on top or a central tower, two storage tanks containing molten salt as a storage and heat transfer fluid, and a power block built of a steam generator, a turbine, and a condenser.

longer periods of low radiation. Since steam turbines can operate economically only above a certain minimum size, the rated output power of today's solar thermal power plants ranges from 50 to 200 mW. The main difference from traditional steam power plants is the solar field, which provides heat for the steam generator. In order to reach the high temperature required to produce steam, solar radiation needs to be strongly concentrated.

For this, only direct sunlight can be used. A mirror that tracks the path of the Sun focuses it on a focal point or a focal line. The higher the concentration, the higher the temperature that can be reached. According to the laws of thermodynamics [65], higher temperatures increase the efficiency of the power plant process. The higher the efficiency, the smaller the collector area required by the power plant to produce the required power output. The technical challenge in the solar field is to achieve the required optical accuracy and robustness against environmental influences such as wind and temperature fluctuations at the lowest possible cost.

As shown in figure 6.8, the solar energy distribution on the Earth's surface is quite inhomogeneous. As the output of a solar energy converter is almost proportional to its input, the cost of solar energy for both photovoltaics (PV) and concentrating solar power (CSP) is strongly related to the selected site.

6.3.1.2 State of the art and future perspectives

Photovoltaics

The worldwide PV market is dominated by solar modules consisting of individual crystalline Si wafer–based solar cells. The process sequence has proven robust and easily expandable using new process developments in research labs. Thus, evolutionary development benefits from technology improvements and the scaling of global production. Both effects have led to a huge reduction of module costs over the past decades (see figure 6.13) and solar module efficiency are typically around 20%.

Such modules can be applied on rooftops with a peak power ranging from a few kWp for residential or large commercial roofs or can be installed in solar parks with several 100 kWp or even MWp capacity. Note that the largest solar parks today have already GWp capacity. As an example, figure 6.11 shows a solar-powered facade of a laboratory building. The blue solar modules are electrically connected in larger strings, and the produced dc electrical current is transmitted by a dc/ac inverter to the electricity grid with an efficiency above 98%.

To calculate the energy produced by 1 kWp of installed PV per year, several parameters have to be considered, including real operation temperature, solar spectrum, and sunshine hours per year. The latter is the most significant contribution to the energy yield per year and is typically 1000 h under moderate climate conditions and above 2000 h under very sunny conditions. Depending on the installation site and application (rooftop, solar park, etc), the levelized costs of electricity provided by PV can range from below 3 cents/kWh up to 7 cents/kWh [66], making PV already the most cost-efficient solution for a growing number of applications. For the future additional cost reductions are expected. Finally, a short

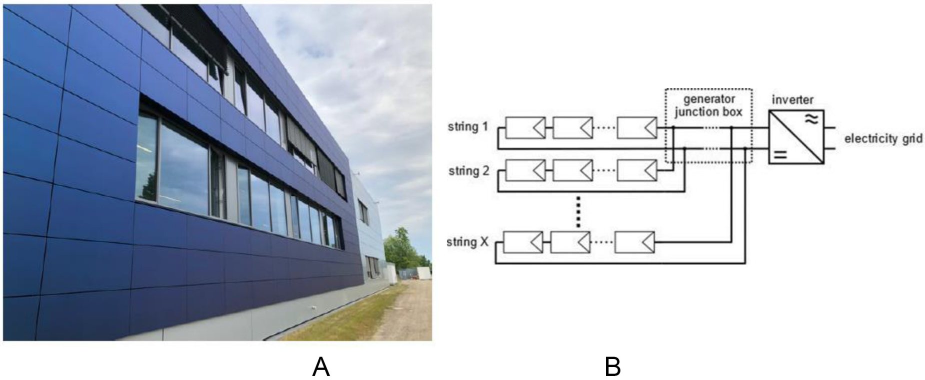


Figure 6.11. (A) Building integrated PV facade as an example of an installation. (B) The electrical layout of the installation. The solar modules are electrically in series connected in strings and bundled in the generator box, and the inverter transfers the dc electricity to the ac internal or external electrical grid.

remark on the energy payback time. With state-of-the-art technologies the amount of energy needed to produce and install a PV rooftop system is delivered back after 1–1.5 years of operation [67], while the operation time of a PV system is longer than 20 years.

PV is still a very young player in the energy sector. As has already been mentioned, significantly higher conversion efficiencies are feasible, already proven by the multijunction concept. Only one highlight example is the unprecedented rise of metal halide perovskites as a new class of PV absorber materials [68], achieving efficiencies well above 20% [64] and close to 30% in tandem configurations with Si [64, 69] within only a few years of systematic research. If the technology can be further developed to module efficiencies surpassing 30% and with low production costs, such tandem PV technologies will pave the way towards further cost reductions.

Solar thermal

In practice, two different basic principles are used to concentrate solar radiation: solar towers and parabolic troughs. In solar tower power plants (figure 6.12(A)), dual-axis tracking mirrors (called heliostats) direct solar radiation to a central receiver mounted on the tower. The heat transfer medium, usually molten salt, water/steam, or air, absorbs energy there and transfers it to the heat storage system and power plant circuit. The surface area of a single heliostat can reach 200 m^2. In commercial power plants, there are thousands of solar towers arranged in a semicircle or circle. Their intense radiation concentration can produce temperatures in excess of 1000 °C at the receiver. In practice, the system operates between 300 °C and 700 °C, depending on the heat transfer medium used [70, 71].

Parabolic trough power plants (figure 6.12(B)) are the most common commercial implementation variant so far. The parabolic mirror trough tracks the Sun uniaxially and focuses the light on an absorber tube running along the focal line.

A B

Figure 6.12. (A) Photo of solar tower plant and (B) parabolic trough plant.

The absorption tube contains a special heat transfer oil. The optically selective coating on the tube absorbs visible light while suppressing heat radiation. The absorption tube is surrounded by a glass envelope tube, similar to a Thermos flask, with a vacuum between the two tubes. This further reduces heat loss. The collector is 7 m wide and 200 m long and uses a hydraulic drive to track the Sun. Today's commercial heat transfer oil allows operating temperatures up to 400 °C.

An important component of solar thermal power plants is the thermal storage system. Two-tank systems with molten salt as the storage medium are most frequently used commercially. In tower power plants, the salt is pumped from the 'cold' tank (at around 300 °C) directly to the solar receiver, where it is heated to over 500 °C and fed to the 'hot' tank. A second circuit takes hot salt as needed and feeds it to the steam generator, from where it is pumped back into the 'cold' tank. In parabolic trough power plants with thermal oil as the heat transfer medium, the salt storage tank is loaded and unloaded indirectly via heat exchangers. As heat can be stored more easily and more economically than electricity, solar thermal power plants can produce solar electricity cost-effectively even after sunset.

Solar thermal power plants are characterized by very low environmental impacts. In particular, greenhouse gas emissions over the entire life cycle are comparatively low. The land requirement roughly corresponds to that of large photovoltaic systems. In the operation of newer power plants, the use of dry cooling significantly reduces water consumption. The effects on the flora and fauna are minor, and only very small amounts of pollutants need to be safely disposed of. In addition, solar thermal power plants have a long service life of up to 40 years [72].

6.3.1.3 Current market situation and perspectives

Photovoltaics

During the past decade PV has developed from a niche technology to a pillar of the electricity supply in a several countries. Global installation reached 700 GWp in 2020 [67].

Expectations for the role of PV in the energy system have changed dramatically. Initially thought to be an economic option only for satellites or remote places, PV is now ubiquitous.

The World Energy Outlook 2020 states that 'solar PV is set for the largest growth of any renewable source. Average generation costs for solar PV have fallen 80% since 2010, and it enjoys support of one kind or another in over 130 countries' [73]. According to the stated Policies Scenario worldwide, solar PV capacity additions will reach 150 GWp per year in 2030. Considering the need to defossilize not only electricity generation, but also the huge industrial sector, transportation, and heating, the global need for PV as a source for electricity but also for 'green hydrogen' is expected to increase significantly. 'In the Sustainable Development Scenario, worldwide solar PV capacity additions reach 280 GWp/year in 2030 and ca. 320 GWp/year in 2040' [73]. The path towards a global installation of 30–70 TWp is discussed in [74] by an international group of PV experts. This scenario is based on very low-cost PV, the enhanced electrification of the energy system, and the ability and strong need to convert power to fuels and chemicals in huge quantities.

Regarding the target for photovoltaic technologies to provide power on the terawatt scale, next-generation PV devices have to provide conversion efficiencies that are as high as possible, provide long-term stability, and are embedded in a circular economy.

A surface area footprint per GWp of solar capacity is estimated to be between 5 and 15 km^2, depending on technology choice. The total required surface area would be up to 400 × 400 km for the targeted 70 TWp. Accordingly, the added value, for example, by being part of the building (BIPV) or of vehicles (VIPV), will help to identify suitable areas. In addition, the use of PV in agriculture needs to be explored and developed. Floating PV plants, using PV on lakes or at sea, will enhance the usable area on our planet.

CSP

In sun-rich countries, CSP technology is suitable to take over the role that today is predominantly filled by fossil fuel power plant, that is, to provide a flexible capacity for periods with low solar or wind input. This capability has been shown by more than 100 commercial solar thermal power plants with a total installed capacity of 6.2 gW. Worldwide, 21 gWh of thermal storage capacity are already installed in CSP power plants, which corresponds to approximately 3 gW of electrical output with an average storage capacity of 7 h [75]. This power can be made available to the power grid as required.

Today, solar power plants are already being planned as an integrated solution to combine PV and CSP power plants at one location, using thermal energy storage to ensure the requirements for security of supply in a cost-effective manner. Thereby, large shares of solar power in the energy system of sunny countries are possible [76].

A significant cost reduction of CSP electricity has been achieved along with its deployment (see figure 6.13). The cost decrease per doubling of the installed capacity (learning rate) is similar to that for PV and higher than that for wind energy; however, since the overall installed capacity is two orders of magnitude smaller than

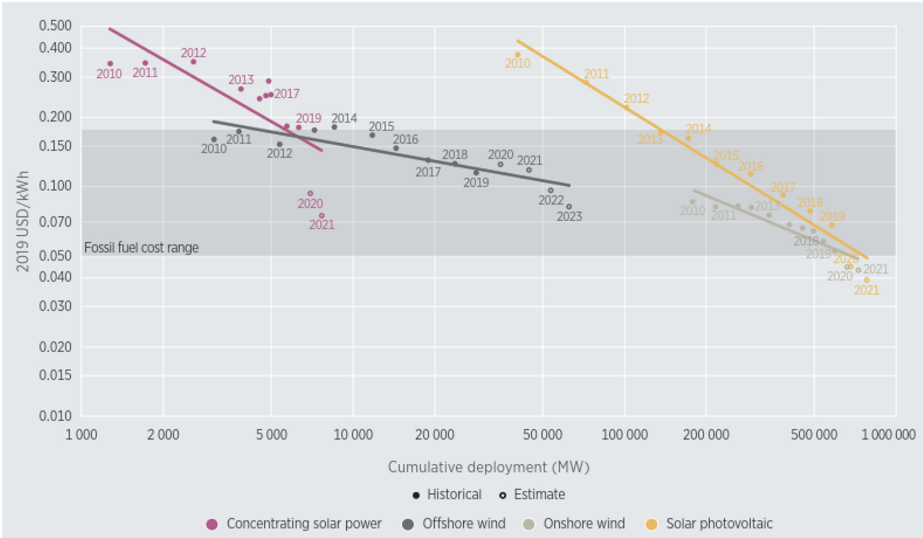


Figure 6.13. Cost reduction of renewable energy technology as a function of its global deployment. Reprinted with permission from [77] IRENA.

that for PV, the levelized cost of energy is higher. With figures approaching 7 US cents per kilowatt-hour in good locations, electricity from solar thermal power plants is already competitive with electricity generated using fossil fuels in places. PV and wind power are offered at lower costs with their integrated thermal storage systems, and solar thermal power plants are the less expensive option for a reliable power supply in times of insufficient input from sunlight and wind, reaching around 5 US cents per kilowatt-hour by 2030 with the help of technical innovations [78].

6.3.1.4 The role of solar technologies for the decarbonization of the global energy system

The threat of climate change requires a fast reduction of anthropogenic greenhouse gas emissions. As most of them are related to the burning of fossil fuels, a massive change in the existing energy system is required. Climate science provided evidence that it is necessary to limit global warming in the next century to less than 2 °C compared to the preindustrial period in order to avoid a catastrophic change of human living conditions on our planet. The cumulative CO_2 emission budget for the energy system that may be acceptable to stay below this target was estimated to be 790 GT [79]. If the emission were kept constant, this budget would be used up in less than 24 years. In the Paris Agreement, the global community set the target to limit global warming to a 1.5 °C increase, based on updated risk analysis. For this goal, less than 7 years are left at today's emission level [80].

Decarbonization of the energy sector to reach the 2 °C target is considered technically feasible if the energy sector can be decarbonized completely by renewable energy and if the share of renewable energy is strongly increased in all other sectors in part through electrification. In addition, a significant increase in energy efficiency in all sectors is required.

To spell this out for solar energy in detail is a bit difficult, as it is not yet clear how the energy will be split among the different renewable resources (in particular wind and solar) and potentially nuclear options and how big the contribution will be in other sectors, for example, through solar fuel production or solar-powered heat pumps.

According to the World Energy Outlook 2020, the 'decisions over the next decade will play a critical role in determining the pathway to 2050' [74]. In all scenarios, renewable energies will play a key role for a sustainable future, and solar energy will become a lead source of our future energy supply. The IEA Net Zero Emissions Scenario [81] predicts a growth of the solar energy supply by a factor of 22 from around 5 EJ solar energy in 2020 to 109 EJ in 2050. Solar energy is expected to provide around 20% of the global energy supply in a decarbonized world. Other scenarios propose even higher shares of solar energy technologies. Thus, it can be considered as one of the most essential technologies to combat climate change.

The installed PV capacity needs to grow from 700 GWp installed in 2020 to many TWp capacity to contribute accordingly. Also, solar thermal power combined with thermal energy storage, today below 10 GW installed capacity, is expected to scale to several 100 GW that will be applied in the sunbelt regions, in particular to supply electricity after sunset. In addition, contributions to cover industrial heat demand is expected by this technology. A possible energy mix progression towards 2050 is depicted in figure 6.14 as a scenario from the IEA [81]. It shows an increasing contribution of solar, wind, biomass, and CCS but also nuclear technologies. This scenario is one of many, with others projecting even larger contribution from solar energy. As scenarios are not predictions but self-consistent options of the future energy system under certain boundary conditions, there is considerable uncertainty

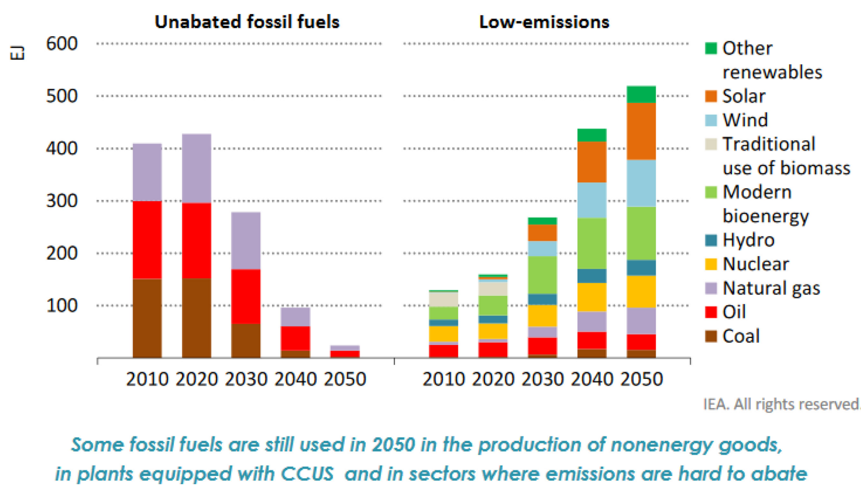


Figure 6.14. Contribution of different technologies o the global energy supply based on the IEA Net-zero Emissions Scenario [IEA 2021]

about the true future energy mix. However, the inevitable contribution of solar energy is undisputed.

Finally, the following arguments support the massive deployment of solar energy over other technology options that may be considered for defossilization:

1. Today's solar energy technologies are mature and already provide electricity at a cost below that of of all other alternatives at more and more sites.
2. Physical principles allow still much higher performance.
3. Learning curves are steep but promise further cost reduction.
4. A combination of PV and solar thermal offers competitive low electricity cost during nonsunshine hour.
5. The cost of electricity storage such as battery storage is declining quickly, allowing low-cost energy storage in countries not suited for CSP.
6. Low-cost solar electricity is also required to produce green H_2 as a key requirement to decarbonize the transport and industry sector.
7. Solar energy technologies have a low environmental impact, and existing technologies do not run into material resource constraints even if scaled as discussed.
8. Production facilities are available and quickly scalable to the required magnitude if the demand is there.

We conclude that a massive exploitation of solar energy is a feasible, low-risk, cost-efficient, and thus no-regret option to achieve the defossilization the global energy system. A continuation of fossil fuel use without considering its climate impact is the only major threat to the rapid deployment of this technology.

6.3.2 Physics of energy: wind energy

Hermann-Josef Wagner[1]
[1]Ruhr-University Bochum, Bochum, Germany

6.3.2.1 General information and facts

Humankind has used wind energy for thousands of years. Even the ancient Persians utilized windmills for pumping water to higher levels. Today, as the leading industrialized countries are debating about the climate change and its dangers, wind energy plays one of the most relevant roles within the renewables. The biggest advantage of wind energy is its clean status. Wind turbines do not pollute the air and do not fill the atmosphere with dangerous gases; this is its strongest characteristic in contrast to conventional energy plants. Furthermore, wind as a natural physical process is available in many regions of the world. Thus, all states of the world need no longer be dependent on the small number of states that own the fossil fuel sources of energy. Unfortunately, because wind is a natural process, one cannot plan and manage its force, velocity, and intensity. This subordinates wind energy systems to conventional power plants, which can produce electricity with a reliable and constant performance. There also exists a location problem. Many people want clean energy, but they do not want a windmill in the neighborhood.

Wind turbines are mostly built in windy land regions and, in recent years, on open sea, so they are called *onshore* or *offshore* turbines. Wind on sea is stronger and more intense than wind on land. Because of the smooth surface of the sea, there is no friction between the wind and objects, so the turbines can make more use of the kinetic wind energy. Onshore turbines are not able to generate as much energy compared to those on sea. The location dilemma and the low acceptance by some citizens are small obstacles which need a solution in future.

There are various types of windmills. Generally, wind turbines are separated into horizontal axis converters and vertical axis converters. Apart from this the number of blades and the rotation velocity define different wind energy converters. Figure 6.15 shows these types.

6.3.2.2 Physics and technical overview

The typical wind turbine is a three-bladed wind energy converter as seen in figure 6.16. These more prevalent than the other types and can be seen in nearly every rural region. A wind turbine generally consists of blades, which are rotated by the kinetic energy of wind and have an aerodynamic profile similar to that of a jet wing. They are attached with bearings and screws to a nacelle, which contains a generator, safety and controlling instruments, and possibly a gearbox. Wind energy systems can contain a gearbox, but they do not have to. Generators with different physical characteristics (multipole equipment) make gearboxes unnecessary. Finally, the plant is held by a steel tower and its foundation.

The functioning of a wind turbine is made possible by physics, especially fluid mechanics. The profile of the blades has a specific design which is also seen in gas and steam turbines, jet wings, and helicopter rotors. The airflow around the blades

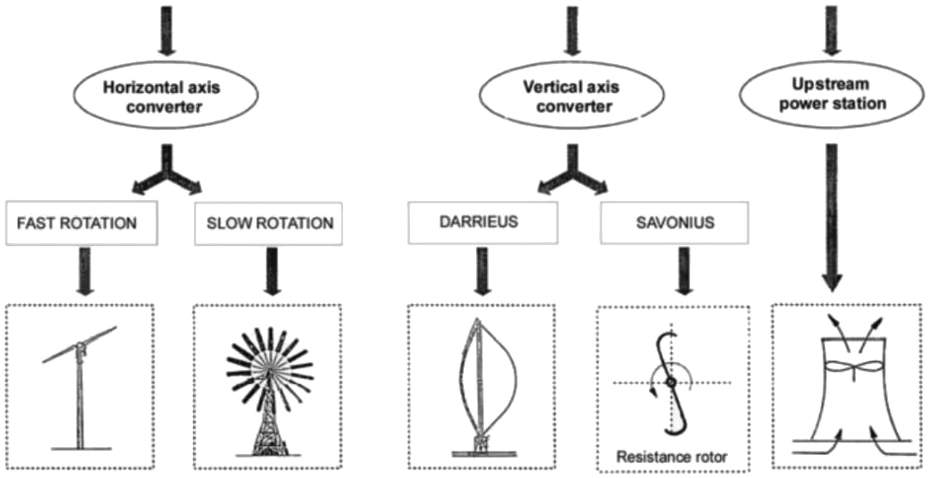


Figure 6.15. Overview of different types of wind energy converters.

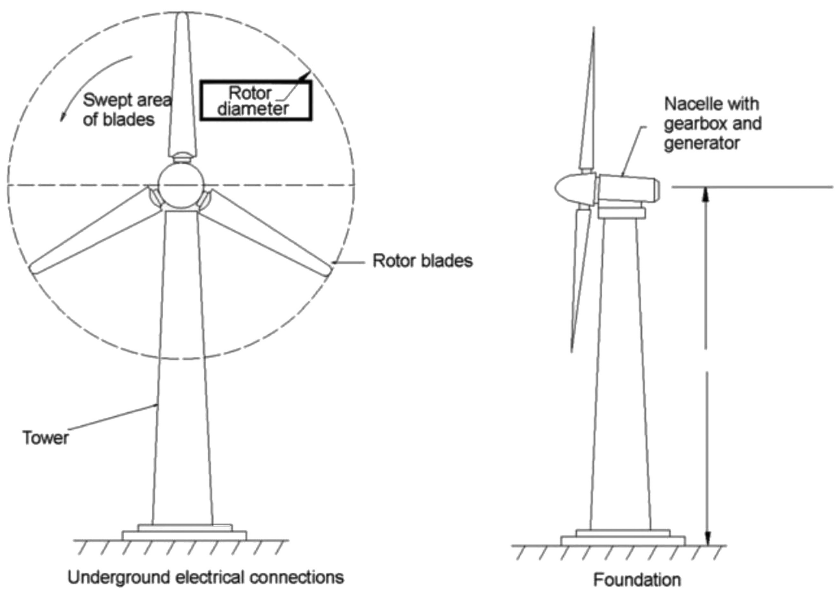


Figure 6.16. Horizontal axis three-blade wind energy converter. A mechanism enables the nacelle to turn in the direction of the wind (rotor speed: 15–25 rpm).

generates a distribution of pressure and different velocities of wind above and under the blade. While low pressure occurs above the blade, high pressure prevails underneath the blade. Based on Bernoulli's fundamental equations, velocity increases with low pressure and decreases with higher pressure. Therefore, the air under the blade flows more slowly than the air above; this results in lift force. Its tangential component lets the blades rotate (figure 6.17). The rotating blades also

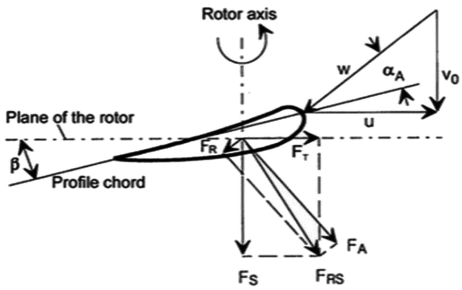


Figure 6.17. Velocities and forces acting on a blade. The tangential component creates the rotation force.

rotate the shaft, which is connected to a gearbox and the generator. The gear ratio changes the rotation speed and allows the generator to build up voltage.

The nacelle of a wind turbine contains a lot of different elements. As was mentioned earlier, gearboxes can be included but do not have to be. Figures 6.18 and 6.19 provide schematic overviews.

With a gearbox: Although the blades are turning slowly, the gearbox can transmit the velocity into higher levels. So the generator, which contains a few pole pairs, is seated on a second shaft and is independent from the rotor velocity. Nevertheless, a gearbox is a complicated mechanical product which needs maintenance. These types can also include a brake, which stops the rotor from turning at stormy winds. Modern plants with gearboxes do not have brakes anymore. They will be stopped by changing the blades angles.

Without a gearbox: These types contain only one shaft, so the velocities of the turning blades and the generator are equal. The loss of many rotating parts is an advantage; possible failures can be reduced. Due to a high number of pole pairs, the generator is still effectively working. Another difference to the model on figure 6.18 is the braking system. The blades can change their angle towards the wind and stop creating the lifting force, which automatically slows down the rotor. A disadvantage of this model is the big diameter of the generator, which is necessary because of the high number of pole pairs in the generator.

Possible electricity generation rates of windmills are as follows:

- Europe: 1500–2500 kWh kW^{-1} installed power
- North Sea: 3800–4500 kWh kW^{-1} installed power

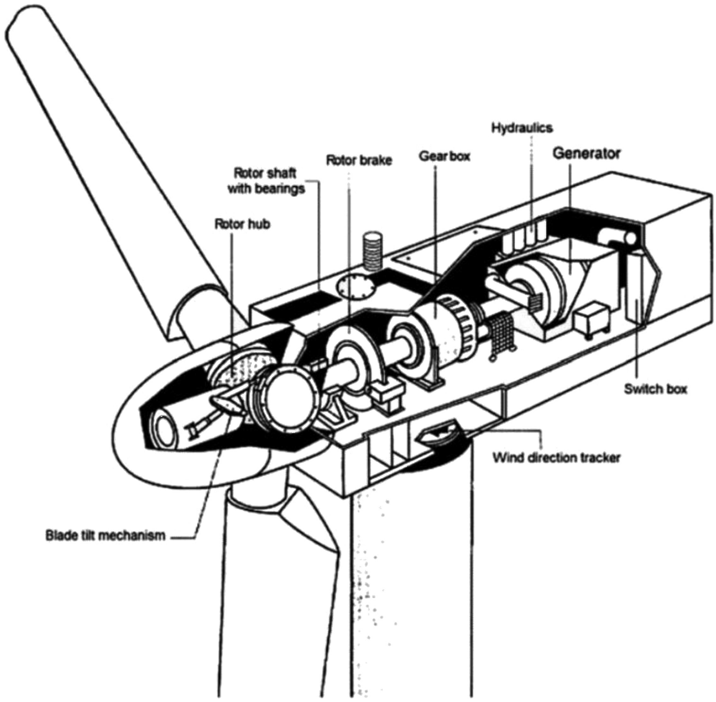


Figure 6.18. Wind turbine with a gearbox.

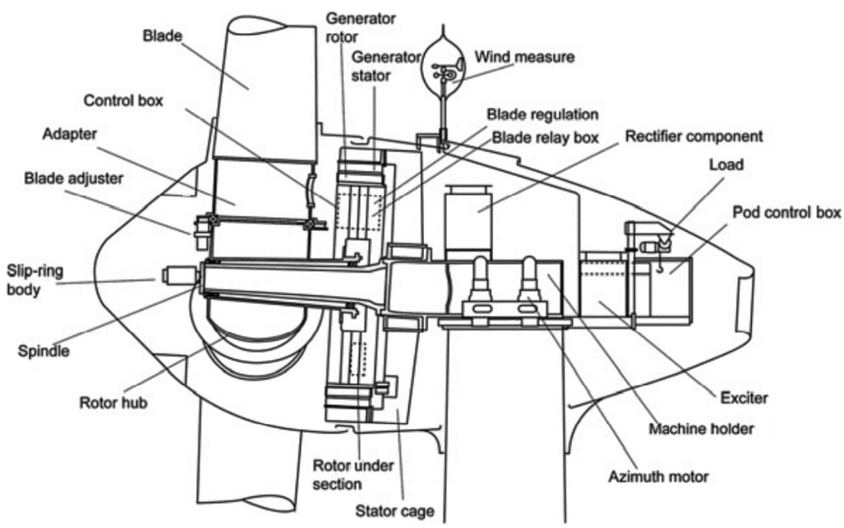


Figure 6.19. Wind turbine without gearbox and details of technical equipment (design of ENERCON Company).

6.3.2.3 The future of wind energy

Today's wind energy technology is still a young form of energy conversion, so scientific and academic research is highly necessary [82]. Because of its clean status and biggest chance to challenge the global climate dilemma, governments all over the world have been helping to expand the wind energy industry through funding. International agreements, such as the Paris Agreement of 2015, obligate all participated countries to follow some rules to keep the climate change low and reversible.

There are numerous new wind energy system variants. One of them is the so-called floating converter. The converter floats like a ship or an oil/gas production facility on the sea. It is fixed with steel cables to the seabed. The biggest advantage is its independence of sea level. Thus, these types can be installed far out in the deep sea (water depth from 50 m to hundreds of meters). This opens opportunities to install large wind parks far out in international sea areas. Of course, maintenance of the turbines and corrosion protection in open sea must be improved.

Furthermore, wind energy converters have a potential to work side by side with gas- and steam-powered plants and hydrogen technology. Unfortunately, engineers still have not found a way of cheaply saving energy created by wind. A battery which can save power on windy days and use it while doldrums could expand the possibilities with wind to higher levels.

One of the main challenges in the next years will be the recycling of old wind energy plants. Like every other mechanical product, wind converters have only a limited lifetime. Especially in the early 2000s a lot of plants were installed, and their deconstruction dates will occur soon. Rotor blades, for example, are made of specific wooden materials such as balsa wood or new materials with higher strength, such as carbon fibre for the blades. The industry still has not found a way to successful recycle these materials in an environmentally friendly way. Researchers are working intensively on this problem.

Ecobalances made for the offshore windpark Alpha Ventus were excellent, but it is necessary to improve them every ten years because of the strong changes in energy supply (due to CO_2 emission reduction activities). The payback time for Alpha Ventus is less than one year.

Important aims which must be fulfilled are:

- More acceptance by citizens living near windmills (through noise reduction, decreased impacts on the local ecology, etc),
- Environmentally friendly processes for recycling of old turbines (especially rotors),
- Connecting wind converters with energy storage modules, such as batteries, for an intermediate time with gas-fired backup power stations until cheap batteries are available.

Finally, the social effects must be calculated too. New wind converter types and their installation support local employment and investment and taxation revenues.

6.3.3 Energy storage

Søren Linderoth[1]
[1]Technical University of Denmark, Lyngby, Denmark

6.3.3.1 History of energy storage

Ever since human beings found out how to utilize energy, we have made use of energy storage. For many years, wood was the way to store energy. To store energy for longer times and make it easier to transport, wood can be converted into charcoal. This is done by pyrolysis, which is a technique in which the wood is converted into charcoal in an oxygen-deficient atmosphere. Nature has also made black coal, which was created millions of years ago. Nature has also made energy-rich gases (natural gas) and oil. Because coal, gas, and oil were created from fossils, they are termed fossil fuels. Fossil fuels were discovered in the 18th century.

When we harvest this stored fossil energy, we may refine and store it further until we need to use the energy. For example, oil is extracted from underground and refined into gasoline that is stored in big tanks; the gasoline is later transported to petrol stations for use by drivers of cars and trucks. We depend on the ability to store the energy.

Fossil fuels are what we use most today worldwide. This fossil energy has advanced the quality of life of humans tremendously. However, now we have to pay the bill. A problem is that large amounts of CO_2, which comes with the burning of the fossil fuels, have been released in a relatively few years. This has caused a significant increase of CO_2 in the atmosphere, which has the effect of warming up the Earth. The atmosphere has become warmer, the polar ice caps are melting, the water level in the oceans is rising, and temperature and weather conditions are changing rapidly all over the world. This is causing severe challenges for many societies. For that reason, significant steps are being taken to replace fossil fuels with energy from renewable sources, such as power from wind turbines and solar cells, solar heating and cooling, use of biomass (wood, straw, etc), and biogas made from waste.

6.3.3.2 Storage of renewable energy

The most efficient way to use renewable energy is to use it directly. However, power from wind and sun fluctuates, so it is not well aligned with our need for energy use. We therefore have to make ourselves independent of when the wind is blowing, and when the sun is shining; that is, we have to find ways to store the energy. Hence, energy storage continues to be a very strong need; in fact it is essential for the transition into a society free from fossil energy supplies. In a sense, we have to do something like what Nature has done for millions of years: make fuels from solar energy and to store the energy, in this case in electrical batteries, as thermal energy, or as mechanical energy such as in flywheels (kinetic energy) or as hydropower (potential energy).

Here, we will discuss some of the important ways to store energy in the future. In order to be useful, the storage technologies must be sufficiently cheap, and the energy must be available in the form we want when we want it. In the same way as energy production must be sustainable, so must energy storage be economically, environmentally, and socially sustainable.

The most important ways to store energy are as follows:

- Electrochemical energy storage
    1. in batteries
    2. by conversion into chemicals (e.g., hydrogen, ammonia, methanol)
- Thermal energy storage
- Potential energy storage, as in dams for hydropower
- Kinetic energy storage, as in flywheels

6.3.3.3 Electrochemical storage

The types of electrochemical energy storage have some common features. They rely on a membrane, also called a separator or an electrolyte, which can conduct only ions and not electrons (or holes), that is, they are solely ionic conductors and not electronic conductors. The electronic conduction must then take place through electronic conductors that can pass through the membrane, leading the electrons from one side of the membrane to the other. The electrodes on the two sides are called the anode and the cathode, depending on whether they are 'charging' or 'decharging'. The key for all is that they consist of an electrolyte (or membrane) and two electrodes. The electrolyte must in all cases conduct only ions and not electrons (figure 6.20). That is an important aspect of electrochemical devices such as batteries

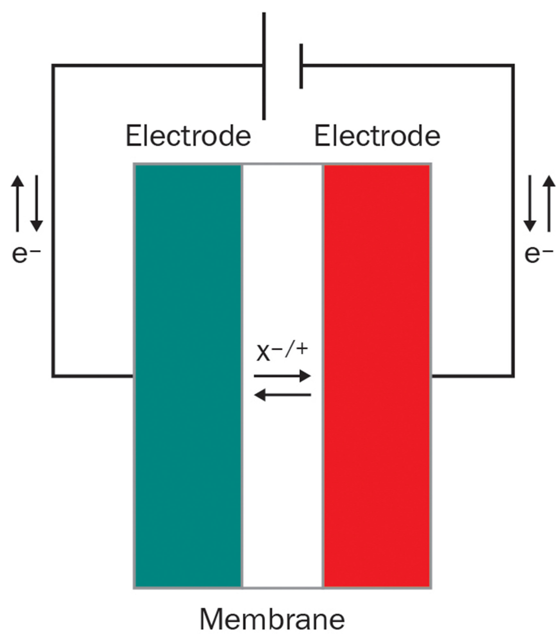


Figure 6.20. Sketch of the heart of electrochemical conversion and storage. The membrane is sandwiched between two electrodes (called anode and the cathode, depending on the direction of the current). The membrane can conduct only ions, not electrons.

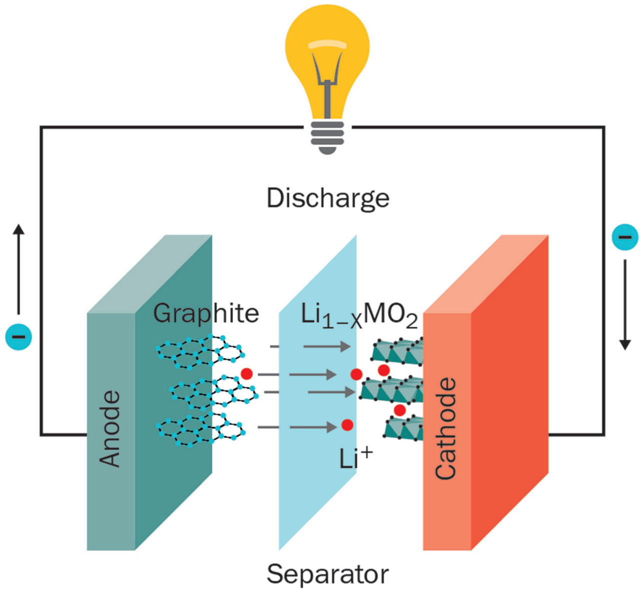


Figure 6.21. Sketch of the function of a Li-ion battery. The 'M' is Mn, Co, Ni, or a mixture of those. When the battery is charged by applying a voltage across the cell, the Li^+ ions are forced to leave the $Li_{1-x}MO_2$ phase through the separator and intercalate instead in the graphite. The separator conducts Li^+ ions. The figure shows the function in the discharging mode, where Li^+ ions intercalated in carbon are transferred through the separator to the cathode, where they are incorporated in the $Li_{1-x}MO_2$ phase (thereby decreasing the x, the deficiency of Li in the phase).

and electrolysers. Examples of ions that can conduct in the electrolyte are H^+, OH^-, O^{2-}, Li^+, and Na^+.

6.3.3.4 Electrochemical storage in batteries

Batteries are electrochemical storage units that contain the stored energy. Many types of battery are portable, like the lead-acid batteries in vehicles and the Li-ion batteries in mobile phones (figure 6.21). Lead-acid batteries are still the most produced and used battery type, especially in cars and trucks. However, the use of Li-ion batteries is increasing fast with the implementation of battery-powered vehicles.

Flow batteries are a quite different type of battery (figure 6.22). The principle is the same: the ions are stored in phases in the anode and cathode, and only ions pass through a membrane. In flow batteries the anodes and cathodes are in a liquid form. The electrodes are the current conductors, and electrons move in and out of the system during charging and decharging. The ions are not stored in the electrodes, as is the case in Li-ion batteries, but are stored in liquids, called anolytes and catholytes.

Yet another type of flow batteries, which will be able to store large amounts of energy, are called gas flow batteries. For gas flow batteries, the liquids in the traditional flow batteries are replaced by gases such as CO_2 and methane (CH_4).

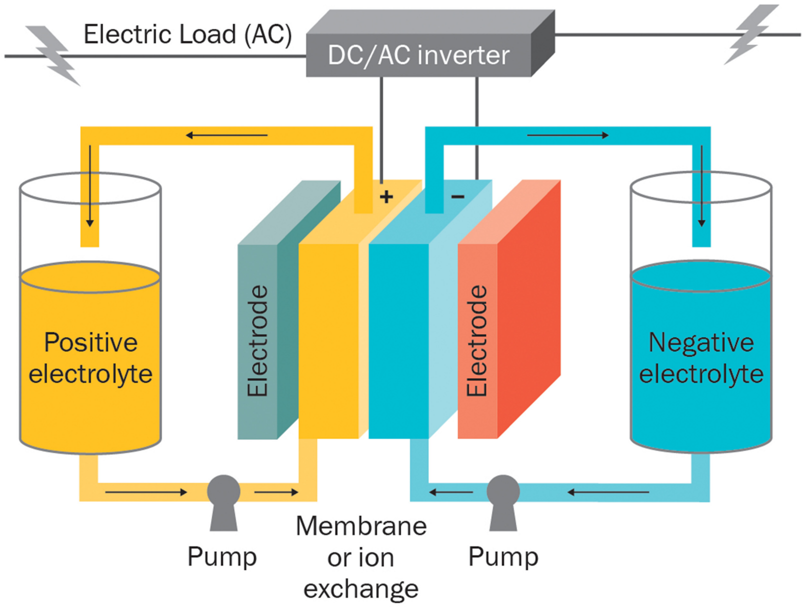


Figure 6.22. Example of the most developed type of flow batteries, termed vanadium flow batteries. Here the valencies are changing in the anolytes and catholytes by a redox cycle. The membrane is a proton-conducting membrane. Anolytes and catholytes made from organic-based materials are under development.

6.3.3.5 Electrochemical energy storage by conversion into chemicals

Electrochemical conversion of power into chemicals can be done; the simplest is hydrogen (H_2), but conversion to ammonia (NH_3), methanol (CH_3OH), methane (CH_4), and other chemicals is also possible. By being converted into chemicals, energy can be stored for a long time and can be transported in various ways and used where electricity is not easy or possible. This could be, for example, in heavy transport, such as long-range flights, or in long-range marine applications.

The core of electrochemical conversion of power is electrolysis. As in batteries, the heart of an electrolyser consists of a membrane/electrolyte/separator sandwiched between two electrodes, the anode and the cathode. The prime difference is that an electrolyser can continue to produce products (charging) as long as it fed with power and the elements for the production. In the simplest case, this is water (H_2O), and the products are hydrogen (H_2) and oxygen (O_2).

In fact, electrolysers have been used for about two centuries, primarily for the production of ammonia for agricultural use as fertilizer. Electrolysers have also been used in industry for production of chloride from NaCl. After many years of slow progress in the development of electrolysers, the new trend of move away from fossil fuels and instead use renewable energy for all energy use has revitalized greatly the need for electrolysis for energy conversion and storage.

The first generation of electrolysers was the alkaline electrolyser with potassium hydroxide and asbestos as the separator. OH^- is the moving ion that is transported from one side of the separator to the other. Asbestos has been replaced by a ceramic–polymer matrix or porous ceramic separator [83]. KOH is very corrosive, and therefore the separator material must be corrosion resistant for use with concentrated KOH. The electrodes are based on Ni [84]. In the beginning, the alkaline electrolysers operated at ambient pressure, but more recent electrolyser have been pressurized (up to about 30 bar), and lately alkaline electrolysers have been developed that operate at above 100 °C.

Another type of electrolyser system that was developed in second half of 20th century is based on a polymer which can conduct protons (H^+). These systems are called proton exchange membrane (PEM) electrolysers. PEM electrolysers need platinum and iridium for the electrodes, which makes them more expensive and therefore more difficult to bring to very large utilization. PEM electrolysers are able to electrochemically pump the hydrogen to high pressures, which is useful for storage of the hydrogen in pressurized tanks, such as is the case in the use of hydrogen in electrical vehicles using hydrogen together with fuels cells for powering the electrical motor.

A third type of electrolyser that is under development and deployment is based on a ceramic electrolyte which conducts oxygen ions (O^{2-}). The electrolyte is based on yttrium-doped zirconium; the doping makes the ceramic a good oxygen ion conductor but at elevated temperatures (700–1000 °C). The electrodes are ceramic or metallic–ceramic (cermets). This type of electrolyser is termed a solid oxide electrolyser cell (SOEC) and is so far the most efficient of types of electrolyser. Other benefits of SOEC is that it also can electrolyse CO_2 and make CO [85]. By electrolyzing simultaneously water steam and CO_2 gas, it is possible to produce a mixture of H_2 + CO, which is called a syngas in industry; from this, one can produce, for example, methanol and methane. Another advantage of SOEC is that the same unit can work the other way around, that is, as a fuel cell (SOFC), and thereby works similarly to a battery [86].

All types of electrolysers are being developed for upscaling and cheap production of hydrogen and other chemicals (figure 6.23).

6.3.3.6 Thermal energy storage

Thermal energy storage is typically the storage of hot water in tanks on the top of a roof or in very large reservoirs of hot water. This type of energy storage is called sensible heat storage. Another type of so-called sensible heat storage is by heating rocks to high temperatures. Here the crushed rocks are heated to around 600–800 °C by blowing high-temperature air through the system. The stored energy may then, preferably within hours or few days, be used to produce high-temperature heat for industrial use, for electricity production, or simply for heating houses. The energy may also be stored as high- or low-temperatures reservoirs in rocks [87] for higher-efficiency electrical production later on.

Molten salts are also used for heat storage. In this case the storage relies on latent heat related to the phase change between the liquid form and the solid form of the

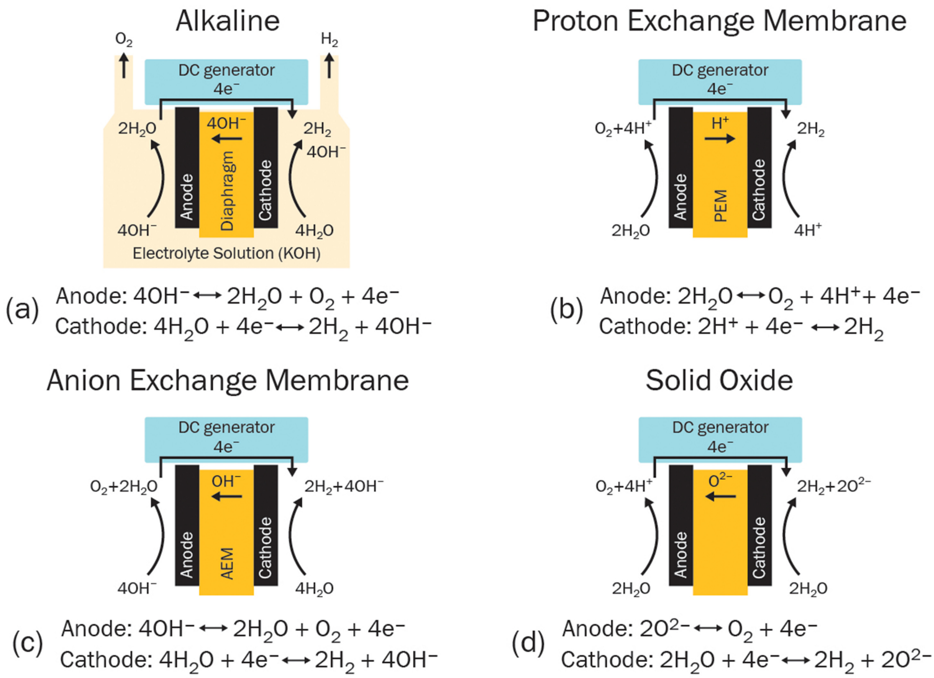


Figure 6.23. Different types of commercially available electrolysis technologies.

material. The temperature range for molten salts is typically from 200 °C to several hundred degrees Celsius. Other phase change materials [88] include some polymers that work at around room temperature and some salts dissolved in water that operates between 20 °C and 100 °C.

Another type of heat storage is thermochemical heat storage. Here heat is generated when two elements react chemically. This could be hydrogen reacting with a metal, making a metal hydride and releasing reaction heat. The system is recharged by heating and releasing the hydrogen from the hydride. Another example of thermochemical heat storage is in salts, where ammonia (NH_3) reacts with the salt structure, forming a salt–ammonia compound. Again heat is released, here at a temperature dictated by the composition of the salt.

6.3.3.7 Energy stored as potential energy

When energy is stored as potential energy, the force of gravity is utilized for the energy storage. The stored element is typically water, and when released, the potential energy converts into kinetic energy, which then drives a turbine, which produces electricity. This type of energy storage is typically used in countries with mountains, where the water is kept behind walls until the water is allowed to flow to

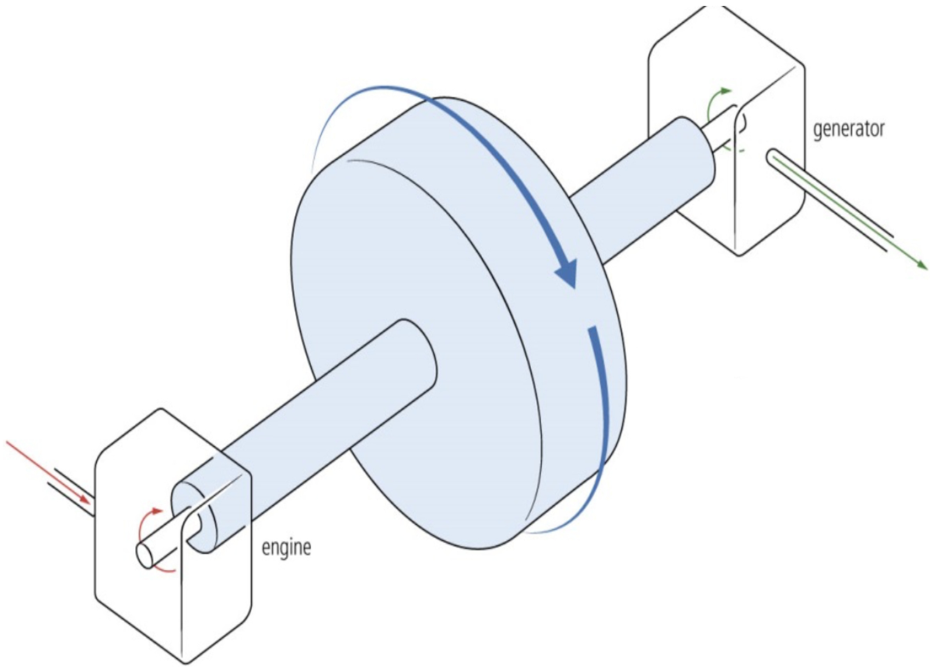


Figure 6.24. Energy can be stored as kinetic energy in a flywheel.

lower heights through turbines. The type of energy that is produced is termed hydropower.

6.3.3.8 Energy stored as kinetic energy

Energy can also be stored as kinetic energy in a flywheel. Here the amount of energy depends simply on the weight of the medium and the speed of rotation. The medium can be glass, metal, or concrete, for example. The stored energy relies on high-speed rotation, and therefore the medium must be able to withstand very high forces that comes with rotation. In order to be able to run at high speeds, the flywheel is typically in vacuum, and magnets are used to lift the system and keep it in place (figure 6.24).

6.3.3.9 Challenges and opportunities in Horizon Europe and beyond

All the energy storage technologies mentioned earlier in this section exist and are used. Therefore, it is sometimes said 'that the technologies exist'. Well, this is correct, just as electrical batteries existed 150 years ago when the first battery-powered car was on the roads and wind turbines have existed for centuries. However, neither the batteries or the wind turbines would be of any use today without the immense development of much more efficient batteries with high energy densities and power densities and the development of much more efficient wind turbines for electricity production. The present industrial revolution using mobile phones and production of cheaper renewable energy would otherwise not have been possible. This is also the case for energy storage. Energy storage has existed and been utilized for many centuries, starting primarily with the use of wood and charcoal and continuing nowadays with batteries and hydrogen. In between, we have utilized the energy storage created by Nature in the form of oil, coal, and natural gas. Because Nature has provided this energy storage, it is cheap. The great challenge now is to reinvent energy storage in many forms that enable cheap energy storage and conversion of renewable energy from solar, wind, and biomass. To make sustainable energy storage—that is the challenge we must meet.

6.3.3.10 Electrochemical energy storage in batteries

The most used type of battery is still lead-based batteries. They are the type of batteries used in traditional cars. These batteries are rather cheap, but the energy density per volume and weight is rather low. In addition, they contain lead, which is not good for the environment and health. On the bright side, the recycling of lead is probably around 99%, although the remaining 1% is not recycled.

To make the transport sector such as cars, buses, and trains, more able to use electrical batteries, high-energy density batteries have to be developed. In 2019 the Nobel Prize in Chemistry was given to three scientists who made it possible to develop the kind of Li-ion batteries we use in our cell phones and in battery-powered vehicles and trains today [89]. Without this breakthrough about 30 years ago, we would not be where we are today with convenient mobile phones and battery-powered vehicles. The development of today's batteries took at least 30 years, and much more has to be discovered and developed to bring the costs down, to make the batteries altogether sustainable, and to make them sufficiently safe.

Li-based batteries contains metals such as Li, Co and Ni. The mining of these elements is a problem. In general, much energy and water are used to extract the metals. Recycling of the materials in Li-ion batteries must come to a much higher level than it is today. Replacing or minimizing the Co could be one aim, and recycling in general is another important aim for Li-based batteries.

The electrolyte in Li-based batteries is a liquid polymer, which can cause fire and therefore can be a risk if used. Other Li-based batteries are based on a solid electrolyte, for example, one made of a ceramic. Such so-called solid state batteries are under development but will need much more development. Some of the challenges for this type of batteries can be dendrite formation [90] in the electrolyte, such that a short-circuit can happen, thereby destroying the battery and causing damage and accidents. Other challenges are the loss of electrical contact between the electrolyte and the electrodes and the reliance on Li metals as one electrode in current solid state batteries. This is greatly beneficial for the energy density, because the volume and the weight of the electrode are much reduced. However, Li metal reacts preferably with oxygen and can become flammable. All of these issues are likely to be solvable but need much effort.

Batteries based on Na^+ as the moving ion are also in use. One type of Na-type batteries is Na-S batteries. They operate at elevated temperatures because they use molten Na as one electrode. The electrolyte is a solid ceramic, beta-alumina (β-alumina/Al_2O_3). Beta-alumina is a good conductor of Na^+ above 250 °C but a poor conductor of electrons, thereby fulfilling the requirement for the electrolyte. Pure Na presents a hazard, because it spontaneously burns in contact with air and moisture, as is the case for Li in Li-metal batteries. Therefore, the system must be protected from water and oxidizing atmospheres. The Na reacts with S on the other electrode, creating Na_2S_4. On charging, this reverses. Na–S batteries have a high energy density and have been considered for use in vehicles. However, the high-temperature operation makes it somewhat difficult, and they are currently used for stationary applications. One challenge is that Na is quite corrosive.

Batteries based on ions such as Al^{3+}, Mg^{2+}, and Zn^+ are under development for rechargeable batteries [91]. Zn–air batteries are well known and used today as primary batteries, that is, used only once, in hearing aids. Magnesium- and aluminum-based batteries still needs much development to be of possible use.

Flow batteries based on organic materials could be very beneficial due to possible nice environmental properties. The first organic-based flow batteries have been demonstrated, but developing more stable organic materials that do not degrade too fast is one challenge. Another is to make the production of the materials cost effective, which will be a requirement for widespread use. Research projects are ongoing but will probably require several years of research and development. For example, flow batteries based on iron compounds are also under development. The materials for these batteries are easily abundant.

6.3.3.11 Electrochemical energy storage by conversion into chemicals

For electrochemical energy storage to be sufficiently sustainable, the cost of renewable energy must be low; therefore, it relies on making energy production from solar and wind even cheaper than it is today. There exist various types of electrolysers, whose cost must be lowered. This can happen in several ways. One is simply by upscaling, which, similarly to, for example, the rapid lowering of costs of Li-ion batteries [92] due to extensive use in battery-powered vehicles, will drive the costs down. However, the electrolysers also need to be improved, and they need to be much more sustainable. The different types of electrolysers systems have different challenge and improvements to cope with. Common challenges are to bring the total cost down for the production of the chemicals and to be able to work well under dynamic conditions, as the input from solar and wind will fluctuate.

Alkaline electrolyser systems should preferably become more efficient in the production of hydrogen. Today, much energy input is lost as heat. Also, the areal need for alkaline electrolysers should preferably be reduced, that is, the power density should be increased. This can be reached by, for example, redesigning the alkaline electrolysers by using new materials that allows the separator to become thinner, and the electrode to be more efficient. New materials and concepts to allow operation at elevated temperatures would improve the power density per volume.

PEM electrolysers would greatly benefit from lowering or replacing the need for platinum and iridium in the electrodes. The membranes of the present PEM electrolysers cannot withstand temperatures above 100 °C. Polymer membranes that can operate well and long-term at higher temperatures could bring higher efficiencies. A challenge in enabling such an increase of operating temperature could be that the very fine-grained (nanoscale) Pt and Ir agglomerate and make the electrode be less efficient and durable.

SOEC operates today at 700–1000 °C. They are very efficient, but the durability of the materials has to be improved. The ceramic materials are brittle, which limits the size of the cells. Today cells are typically around 100 cm^2. To bring down the costs, the cell areas should be larger without causing frequent failures during operation. In general, the durability during operation must be improved by optimizing the materials and the operation. The stack concepts should also be revisited and improved. The most efficient use of SOEC comes with the synergy of synthesis of chemicals, such as methanol. The interplay between these units must also be optimized and developed further.

6.3.3.12 Thermal energy storage

The technologies for storage of hot water in small and large quantities are relatively mature. Thermal energy storage in rocks, in phase change materials, and by thermochemical reactions have virtues for much higher energy densities and for uses other than heating of houses. However, they all need to be reduced in price, and some need to be developed further and demonstrated for up-scaling for large deployment.

The use of phase change materials, such a molten salts, for energy storage is well proven but is waiting for a breakthrough. Corrosion of molten salts is one issue; another challenge is avoiding solidification of the molten salt, which can cause severe problems due to expansion of the medium, causing cracks and faults. In general, cost is an issue.

Thermochemical energy storage has the potential for high thermal energy density, and it can be completely without energy loss, as the heat is realized only upon reaction, which can be controlled. Thermochemical units have not much been demonstrated, and there is a need for further developments to happen. The salt-ammonia system is probably the most energy-dense system. Here, one challenge can be the ammonia, which is poisonous and therefore can limit the use of this system outside industrial applications. Water or steam may replace the ammonia but with much reduced energy density. The recycling of the system must be proven for many thousands of cycles.

6.3.3.13 Energy stored as potential energy

The potential for the use of more potential energy is already quite limited, though it is widely used in hilly countries. Also, the impact of the environment in the hills or mountains can be bad, as the level of the water alternates greatly over the year. This causes damage to the environment. In some areas, however, there are still great possibilities for use of melting ice as a renewable resource. In Greenland more ice is melting each year as the atmosphere becomes warmer. The melting can create a flow of water, which can be used to run turbines whose power can be used for making power-to-X (hydrogen, ammonia, and more).

6.3.3.14 Energy stored as kinetic energy

The benefit of flywheels is the possible large power density, not the energy density. Flywheels can take up and deliver fast energy, for example, in connection with storing the electricity from the stop and start of electrical trains. Flywheels could also be used in cars and buses for storage of energy from deacceleration and used when accelerating. Flywheels have not yet found widespread use, and one aspect of this can be the cost of the systems. The cost would decrease with more use of flywheels. The high speed of the flywheel can be a risk that has to be dealt with. Improving the stability of the spinning wheel by using even better magnetic systems could be a solution.

6.3.3.15 Conclusions

New means of energy storage have to be developed for the green transition to be possible. Such energy storage technologies must come at the lowest possible costs, both in making the storage units and in their operation. Energy storage is needed both for short-term storage with high power output and long-term storage of large amounts of energy. Some types of energy storage, though not all, have been discussed here, and new technologies will appear in the future. However, it is quite certain that electrochemical energy storage in the form of batteries of various kinds is a must, as is electrochemical storage in chemicals provide by electrolysis and synthesis into chemicals and fuels, such as hydrogen, ammonia, and methanol. Thermal energy storage will also be of importance, especially in places with a need for heating or cooling.

In all cases, further research and development, demonstrations, and commercialization are needed, and we have to do this rather quickly, as the goal for many countries, regions, and companies is to be CO_2 neutral by 2045–50. This is in a short amount of time when we consider that many technologies have to be developed to a much higher level and costs have to come down drastically so that the transition from dependence on fossil fuels can be replaced by energy supply coming from renewables sources. For this we ned many good brains and much work.

6.3.4 Fusion energy development

Alberto Loarte [1]
[1]ITER Organization, Saint-Paul-lez-Durance, France

6.3.4.1 Introduction to fusion energy development and present status

The physical processes that lead to the production of energy in the Sun and stars were understood in the first third of the 20th century [93, 94]. Since then, there has been a continuous effort to make use of this understanding to develop an energy source for humankind based on the same principles. The advantages of mastering the fusion processes as an energy source are tremendous, including the wide availability of fusion fuels, no production of greenhouse effect gases, and no long-lived radioactive waste. However, the scientific and engineering challenges to realize this energy source are also considerable, as was determined in the initial research by the mid-20th century.

To overcome such challenges, researchers soon realized that concerted scientific and technical efforts were required. These were undertaken at the national and supranational levels (e.g., the European Union) with a strong international collaborative attitude from the start. The pinnacle of these collaborative efforts is the first experimental fusion reactor (ITER), now under construction in France by a consortium of international members (China, the EU, India, Japan, the Republic of Korea, the Russian Federation, and the United States). ITER's objective is to demonstrate the scientific and technological feasibility of fusion energy as a sustainable source for humankind, and experiments to address this objective will begin in the mid-2030s [95].

The most effective way to produce fusion energy on is through the reaction of two heavy forms (isotopes) of hydrogen: deuterium and tritium. Most of the deuterium was created in the early stages of the formation of the Universe and is thus widely available (typically, about 0.016% of the hydrogen in seawater is deuterium). Tritium, on the other hand, is an unstable form of hydrogen that decays radioactively in 12.5 years and thus needs to be produced. An effective way to produce tritium for fusion energy production is by neutron irradiation of lithium, which is abundant in the Earth's crust. The nuclear reaction of deuterium (D) and tritium (T) produces helium (He) and one neutron:

$$\mathrm{D} + \mathrm{T} \rightarrow \mathrm{He} + n + \text{Kinetic Energy} \; (\sim 3 \times 10^{-12}\ \text{Joules}) \tag{6.7}$$

Since the mass of deuterium and tritium is larger than that of helium plus the neutron, the missing mass is converted to kinetic energy of the helium nucleus and neutron according to Einstein's relativity law for conservation of mass and energy ($E = mc^2$). Because the neutron is four times lighter than the helium nuclei, it carries most of the energy produced in the reaction (75% of the total). This neutron kinetic energy can be transformed into electricity by slowing down the neutrons and heating up water or other fluids in a way similar to that used in present gas, coal, and nuclear fission power stations.

There are, however, two key differences between fusion-based reactors and present fission-based reactors that make the fusion option sustainable in the long term. The first one is that the product of the reaction is helium, which is nonradioactive. The second is that the total kinetic energy gained by the neutrons in the fusion reaction, which ultimately is transformed into electricity, is much larger than that for fission for the same mass of fuel converted into energy. In fact, the electricity consumption needs of one person for 30 years can be satisfied with the deuterium contained in the water in a bathtub and the lithium contained in the battery of a laptop computer when fused in a reactor. The widespread and long-term availability of lithium and deuterium could support the supply of fusion energy to humankind for thousands of years.

To get the deuterium and tritium nuclei to fuse, it is necessary to overcome the electrostatic repulsion force, since both nuclei are positively charged and they repel each other. When the nuclei approach sufficiently close to each other, the nuclear fusion process can effectively take place and ensures high production of fusion energy. This requires deuterium and tritium nuclei to collide at high velocities; this is realized by heating the deuterium and tritium mix to very high temperatures (hundreds of millions of kelvins) (figure 6.25); at these high temperatures the thermal agitation of deuterium and tritium ensures that a sufficient fraction of them have enough velocity for the fusion reaction to take place. In addition, the collisions between deuterium and tritium nuclei should be frequent enough for copious power (i.e., energy per unit of time) production to take place; this implies that the density of deuterium and tritium should be sufficiently high. These density and temperature requirements are met in the cores of the stars and will be met in fusion reactors; the differences are that the stars fuse protium (the lightest type of hydrogen) and not deuterium and tritium, with the Sun having a typical temperature of 10 000 000 K in its core compared to ~200 000 000 K in a fusion reactor. In these high-temperature conditions, the hydrogen gas is ionized and electrons are not bound anymore by electrostatic forces to the hydrogenic nuclei. This state of matter is called plasma; it is the most abundant state of matter in the Universe (although not on the Earth's surface) and can exist over a wide range of temperatures and densities (figure 6.26).

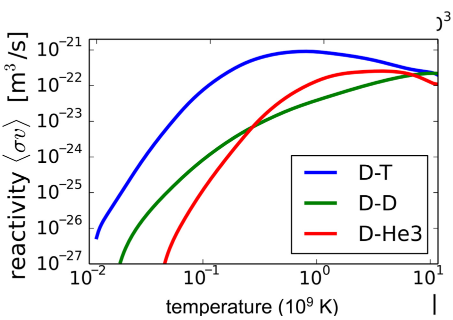


Figure 6.25. Reactivity of three fusion reactions versus temperature showing the higher reactivity of the deuterium–tritium reaction compared to the others and that it reaches its maximum value at temperatures of few hundred million kelvins. (Source: https://en.wikipedia.org/wiki/Nuclear_fusion. Credit: Dstrozzi)

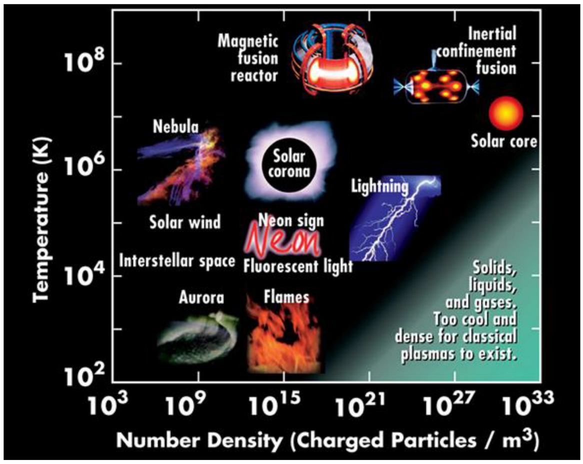


Figure 6.26. Typical temperatures and densities of plasmas on the Earth and in the Universe compared to those to be achieved in fusion reactors. (Source: https://www.cpepphysics.org/fusion.html. Credit Contemporary Physics Education Project (CPEP), used by permission.)

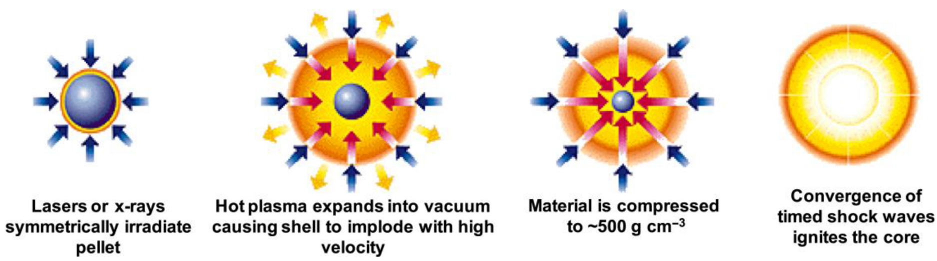


Figure 6.27. Diagram of the dynamics of fusion energy production by inertial confinement by direct. (Source: https://www.lanl.gov/projects/dense-plasma-theory/background/dense-laboratory-plasmas.php. Credit Los Alamos National Laboratory.)

Such hot plasmas lose heat by conduction and convection, and they naturally expand because of their pressure in a similar way to hot air in a balloon. To sustain fusion power production in these plasmas, it is necessary to keep the plasmas hot and with sufficient pressure, since expansion and heat losses would decrease the plasma temperature and density and stop the fusion reactions. Stars achieve these goals thanks to their huge dimensions, which slow down heat losses, and to their mass, which provides the gravitational force to compensate the expansion of the plasma. To achieve the same goals as in the stars and achieve fusion power production on Earth, other physical processes are required; two approaches have been developed the so-called inertial confinement fusion and magnetic confinement fusion.

In inertial confinement fusion, a small solid deuterium–tritium spherical shell is irradiated by high-power-density light, which heats and compresses the shell to high densities and temperatures, leading to production of fusion power in short bursts (figure 6.27).

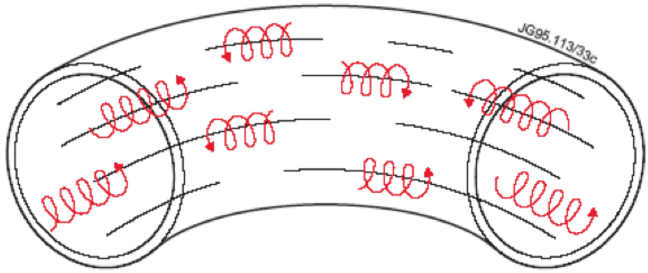


Figure 6.28. Trajectories of ionized charges particles (red) along magnetic field lines (black) in a section of a torus. (Source: https://www.euro-fusion.org. Credit EUROfusion consortium.)

Magnetic confinement fusion takes advantage of the physics processes that determine the movement of charged particles in magnetic fields. A charged particle (with charge q) with velocity $\vec{v}$ in a magnetic field $\vec{B}$ is subject to a force (so-called Lorentz force) given by

$$\vec{F} = q \quad \vec{v} \times \vec{B} \tag{6.8}$$

The Lorentz force ties the movement of particles to the lines of the magnetic fields (figure 6.28). For hot plasmas this decreases the heat losses in the direction perpendicular to the field (note that $\vec{F}$ in equation (6.8) is always perpendicular to $\vec{B}$) but has no impact in the parallel direction to the field. To avoid losses in this direction, it is necessary to use magnetic field lines that close on themselves, forming a torus (or doughnut) with a magnetic field (or toroidal field) around the axis of symmetry of the torus. This is not sufficient to ensure that the plasma is in equilibrium, and another component to the field in the short direction around the torus (or poloidal direction) must be added. The most successful magnetic field configurations for fusion development are the tokamak and the stellarator (figure 6.29). In the tokamak, the magnetic field is produced by electric currents circulating in external coils to the plasma and within the plasma itself; in the stellarator the magnetic field is produced by external coils only. Since the generation of these magnetic fields requires electric currents and thus power, which would decrease the net power production by the reactor, it is important to optimize the magnetic fields so that they provide the required thermal insulation and compression force to maintain the fusing deuterium–tritium plasma with the minimum power used for their generation. Typically, the magnetic fields to be applied to the plasma are in the multi-Tesla range (or ~100 000 times the Earth's magnetic field) and are created by coils surrounding the plasma in which large electric currents circulate, typically in the multimegaampere range (or several million times the electric currents used at home).

For both tokamak and stellarators it is necessary to heat the plasma to sufficiently high temperatures so that the fusion reaction is effective. This is done by the injection of radiofrequency waves that couple to the movement of electrons and ions in the magnetic field (shown in figure 6.27) and accelerate them (similar to the physics processes to heat food in a microwave oven) or by the injection of high-energy

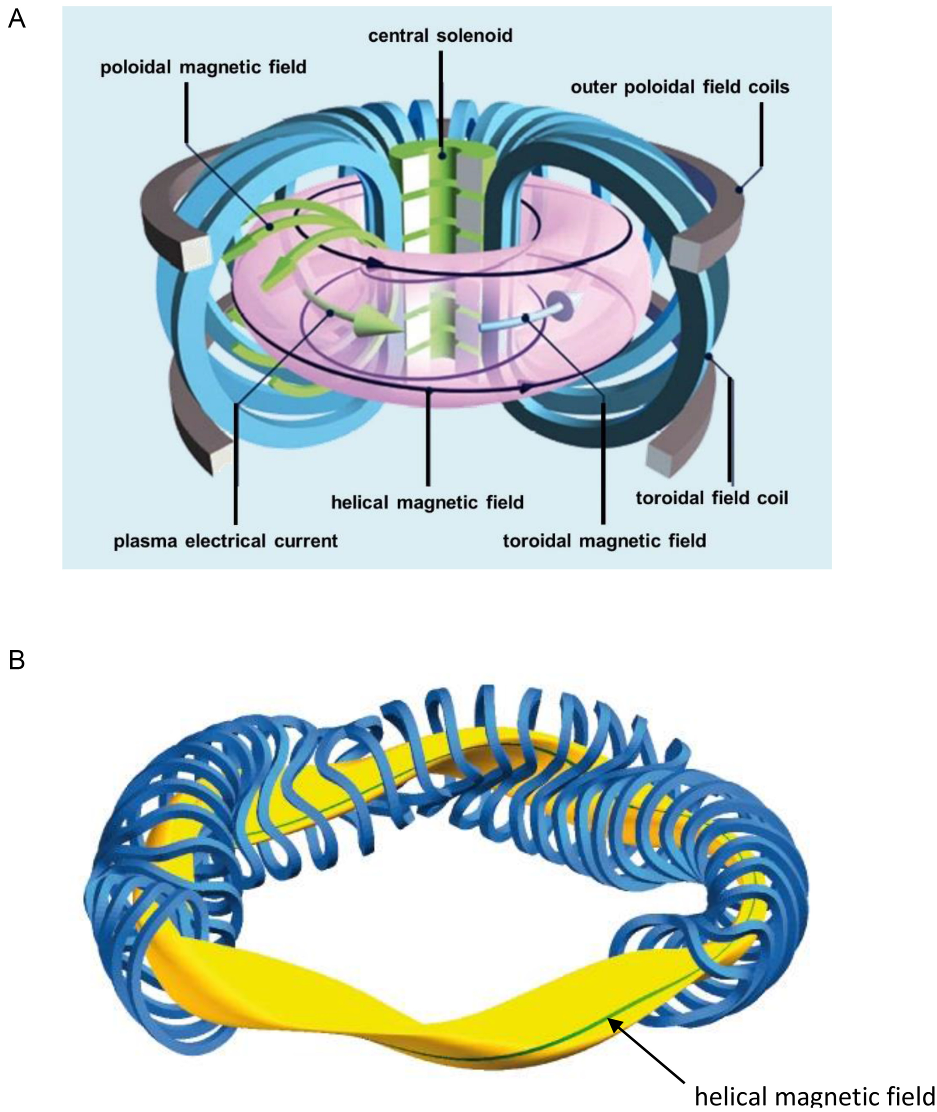


Figure 6.29. Schematic description of the coils and magnetic fields in the tokamak (a) and stellerator (b) magnetic confinement schemes and the resulting helical fields. (Source: https://www.euro-fusion.org, Credit EUROfusion consortium, and http://www.ipp.mpg.de, credit Max-Planck-Institut für Plasmaphysik, copyright IPP.)

hydrogenic neutrals that thermalize in the plasma and increase its temperature (similar to adding a bit of boiling water to cold water to get lukewarm water). Once high temperatures are achieved and the DT fusion reaction starts, the plasma starts to self-heat by the high-energy helium produced in the fusion reaction, which carry 25% of the energy produced in the DT reaction (equation (6.7)). At higher fusion reaction rates, this becomes the dominant source of plasma heating, with other heating sources external to the plasma being used to control the fusion reactivity.

In fact, the thermal insulation provided by the magnetic fields is limited by turbulent processes; therefore, the heat injected by external means and produced by plasma self-heating eventually escapes through the plasma external surface and is deposited on the material surfaces that surround the plasma. Since this thermal insulation is very good, it is possible to maintain temperatures of hundreds of million kelvins in the central part of the plasma, while it is typically a few million kelvins at the plasma external surface. In magnetic fusion reactors, the plasma itself is contained in a toroidally shaped vessel surrounded by the coils that create the magnetic fields. This vessel is clad with a wall of protective elements that are in direct contact with the plasma and ultimately receive the heat and neutrons produced by the fusion process. Neutrons, which carry 75% of the energy produced by fusion, are not electrically charged and are not subject to the Lorentz force (equation (6.8)) and thus distribute themselves uniformly on the protective elements. By contrast, the plasma particles are charged and tied to the magnetic field by the Lorentz force, which makes the plasma losses to the vessel protective elements flow along the magnetic field lines. This means that the heat losses from the plasma are deposited upon a very small fraction of the available surface of the vessel wall (typically a few percent), and this leads to very large power fluxes (flux = total heat/area). Even if only 25% of the energy produced by fusion reactions is deposited in the plasma itself by the energetic helium, the power fluxes deposited on the wall by the plasma can be 100 times larger than those from neutrons in a fusion reactor. Such concentrated plasma fluxes can cause local overheating and erosion of the wall components, which pose specific challenges, as will be discussed in the next section.

The effectiveness of fusion production is characterized by the fusion gain (Q), which measures how much power is produced by the deuterium–tritium plasma compared to that used to heat up the plasma by external means:

$$Q = \frac{P_{\mathrm{DT\ fusion}}}{P_{\mathrm{heat}}} = \frac{P_{\mathrm{He}} + P_n}{P_{\mathrm{heat}}} \tag{6.9}$$

where P_{He} and P_n is the power produced by the deuterium–tritium fusion reaction in the form of energetic helium and neutrons in equation (6.7) and P_{heat} is the external power required to heat the plasma. This can be directly related to the product of the plasma density (n), temperature (T), and the so-called energy confinement time (τ_E), which characterizes the time for plasma heat losses (i.e., the characteristic time in which the plasma loses its energy when heating stops). This triple product is commonly used to quantify progress in fusion energy research, since most of the present experimental facilities do not make use of deuterium–tritium plasmas but only deuterium plasmas. In these plasmas the equivalent production of fusion power by deuterium–tritium can be easily quantified by the triple product. Note that actual DT fusion power production in magnetic confinement devices has actually been demonstrated in only two tokamak experiments (TFTR and JET), with JET presently carrying out a new series of DT experiments.

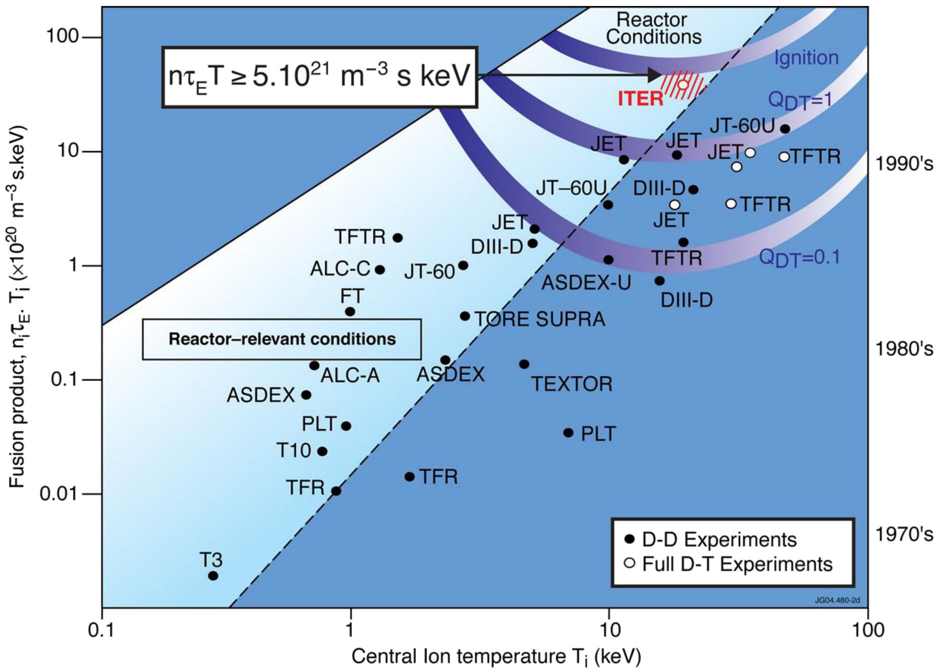


Figure 6.30. Experimentally achieved values of the fusion triple products versus temperature of the plasma hydrogenic ions in kiloelectronvolts (1 keV ≈ 12 000 000 K) in a range of tokamak devices and expected values in ITER. (Source: https://www.euro-fusion.org. Credit EUROfusion consortium.)

The challenge of magnetic confinement fusion energy development is thus the creation and maintenance of high-temperature plasmas with sufficiently high density and low heat losses (e.g., long τ_E) for the fusion reactions to occur and self-heat the plasma for long durations (minutes to hours). This has been the focus of fusion research in experimental facilities, and significant progress has been made, as summarized in figure 6.30. As shown in this figure, deuterium–tritium equivalent fusion power production in present experimental facilities has reached $Q \leqslant 1$ values, meaning that more power is required to heat the plasma up than is produced by fusion processes. Although not shown in this figure, the duration of the high fusion power production phases is restricted to a few seconds at most. Experiments have demonstrated that to achieve and sustain fusion power production over long time scales, it is necessary to understand the physics processes not only of the hot deuterium–tritium plasma but also of those at the outer colder layers of the plasma and in the interaction zone between the plasma and the wall-protective elements. Therefore, although the individual physics processes required to achieve fusion power production have been identified and thoroughly explored in present experimental facilities, their integration in a plasma dominated by self-heating from fusion reactions (i.e., the fast helium produced by deuterium–tritium fusion) remains an open challenge.

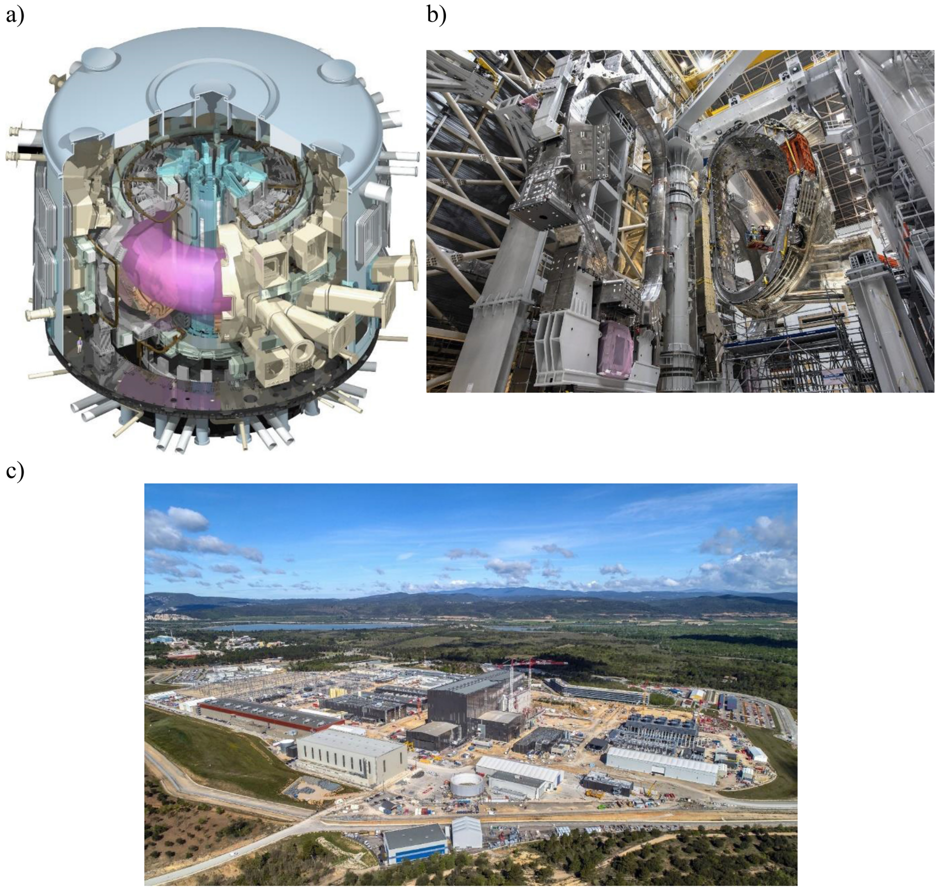


Figure 6.31. (A) Cutaway drawing of the ITER tokamak showing the main components enclosed in the cryostat and an artistic impression of a plasma inside its vessel. (B) First portion of ITER vessel being assembled with two coils that will produced the toroidal magnetic field. (C) Aerial view of the ITER site with the buildings where the tokamak is assembled and will operate (center, in black) and the supporting ancillary plants and buildings spread over 42 hectares. (Source: https://www.iter.org. Credit ITER Organization.)

The demonstration of high Q fusion power production for extended times of minutes up to an hour is the mission of the first generation of experimental fusion reactor devices, chiefly the ITER tokamak that is presently under construction (figure 6.31). The physics basis for the design of the ITER tokamak rests solidly on the physics understanding provided by the present experimental fusion facilities and theoretical developments [96]. Specifically, ITER is designed to demonstrate fusion power production with $Q \geqslant 10$ for durations of 5–8 min and aims to demonstrate $Q \geqslant 5$ for durations of up to 1 h [95]. These goals have been considered for the determination of the physical dimensions of the plasma (with a volume of 840 m^3, i.e., ten times larger than existing fusion facilities), the magnetic fields to be applied, the electric current created in the plasma, and so on, but they also have a direct impact on technological choices for the components of the device. For example, heat dissipation by ohmic heating if the ITER coils were made of copper would be very large; therefore, superconducting materials are used. These superconducting coils have negligible electric resistance and thus are not heated up by the electric currents that circulate in them. For the superconducting coils to reach such conditions requires very low temperatures (a few degrees Kelvin) and thus strong refrigeration (by liquid helium) and very good insulation from surrounding heat sources. This is provided by installing the coils inside a cryostat (a 'Thermos flask' of very large dimensions) and in vacuum.

The size of the ITER tokamak and its ancillary systems, the complexity of the technologies required for the achievement of its goals, and the unprecedented plasma parameters that will be achieved for the first time in a self-heated plasma by deuterium–tritium fusion processes make this experimental facility unique for addressing the challenges ahead in the development of fusion as an energy source. Its international cooperation dimension is also a demonstration of the importance of ITER's goals for humankind.

ITER by itself will be a major first step in the demonstration of fusion energy as an energy source. ITER's scientific exploitation will be accompanied by other, smaller-scale devices that will address specific fusion physics issues in more detail that can be explored in ITER or will explore different approaches to the tokamak concept on which ITER is based. The knowledge acquired in ITER and these accompanying devices will be used for the construction of DEMO [97], the first demonstration fusion power plant with net electricity production from deuterium fusion reactions. In the next section, we discuss the scientific challenges that need to be faced to achieve this goal and the contributions from ITER to these challenges.

6.3.4.2 Challenges for the development of fusion energy: ITER and beyond

As was mentioned previously, the main challenges to be resolved in ITER and future magnetic confinement fusion reactors concern the simultaneous integration and control of the key physics processes that are required to achieve a plasma dominantly heated by deuterium–tritium fusion leading to high-gain fusion power production. To face these challenges, an understanding of the underlying physics processes determining heat and particle losses from the plasma, the stability of the equilibrium of forces between the plasma and the magnetic fields, and the integration of fusion plasmas with wall requirements, to cite the major key challenges, is required. Conceptually, such challenges are similar for the main magnetic confinement schemes (tokamaks and stellarator), although quantitatively they can substantially differ.

Confinement and transport in self-heated fusion plasmas

The conductive and conductive losses from the plasma are determined by turbulent transport processes driven by the variations of plasma parameters across the magnetic fields. The dominant turbulent transport processes that will take place in

ITER plasmas are well established from present experimental results and a well-developed plasma theory/modeling basis. However, the magnitude of the turbulent processes is difficult to predict quantitatively for plasmas in ITER and future fusion reactors, since they will have densities and temperatures that cannot be achieved simultaneously in present experimental devices. The precise magnitude of turbulent transport has a direct impact on the plasma parameters and thus on the self-heating of the plasma by fusion reactions. Since plasma self-heating dominates external heating in ITER and fusion reactors, unlike in plasmas of present experimental devices, a complex feedback loop among plasma parameters, self-heating, and plasma transport will be established. The detailed balance of these processes affects the efficiency in fusion power production (quantified by Q) that will be achievable in the reactor.

It is of particular importance to understand the physics of convective losses (driven by the flow of plasma ions and electrons) in comparison to conductive losses (driven by temperature gradients in the plasma) and optimize their ratio as far as possible. Reducing conductive losses increases the plasma temperature and fusion power production while reducing convective losses; this can have detrimental effects if these become too small. If plasma flows from the core to the periphery are not sufficiently intense, this can cause the accumulation in the plasma of the helium ‘ash’ produced by the fusion reaction and other impurities present in the plasma (resulting from interaction of the plasma with the wall-protective elements), and this can eventually stop the reaction altogether.

Convective transport also plays the essential role of replenishing the burnt deuterium and tritium in the plasma core where fusion reactions take place. If no ‘fresh’ deuterium and tritium are provided to the plasma core, fusion energy production will stop because of lack of fuels. Directly injecting deuterium and tritium into a hot fusion plasma is not possible; due to the high plasma temperatures of fusion plasmas (tens to hundreds of millions of kelvins), any deuterium or tritium atom entering the plasma is immediately ionized and thus tied to the magnetic field. To increase the deuterium and tritium fueling efficiency, these are injected into the plasma at high speed in the form of solid pellets, which first vaporize and then ionize. This allows increasing the penetration of the deuterium and tritium fuels up to several tens of centimeters into the plasma, which is still small compared to the typical dimensions of several meters of a fusion plasma. Therefore, detailed understanding of the convection of ionized deuterium and tritium from the plasma periphery towards the plasma core is key to achieving the replenishment of the burnt fuel; this is a rich physics area, since turbulent transport affects deuterium and tritium ions differently, due to their different masses.

Understanding the physics processes that dominate the feedback loops that will be established in self-heated plasmas and how to use external actuators (electric currents in the plasma, magnetic fields, etc) to optimize them for maximum fusion power production is a major objective of the ITER scientific program. The fusion gain of $Q \geqslant 10$, for which ITER is designed, was specifically chosen to ensure that the physics of self-heated plasmas can be investigated. For a $Q \geqslant 10$ fusion plasma, the internal plasma self-heating by fusion reactions is at least twice that provided by external heating sources. For demonstration fusion power plants with net electricity production after ITER, such as DEMO, the ratio of self-heating to external heating will increase to 5 or more.

Plasma stability control

Magnetic confinement fusion plasmas maintain equilibrium by the magnetic fields creating a force that opposes the expansion of the hot plasma, and this equilibrium may become unstable. These instabilities (so-called magnetohydrodynamic instabilities) directly affect the achievable fusion power production, since this requires high density and temperature in the reactor plasma and thus high pressures (in the range of 5–10 times the atmospheric pressure in the central plasma region). Instabilities in the plasma equilibrium can be local or global; local instabilities lead to a rearrangement of magnetic fields and plasma pressure in a toroidal annulus of the plasma and can cause a moderate decrease of fusion power production, while global instabilities cause the loss of magnetic confinement, cool-down of the plasma, and stopping of the fusion reactions. An example of such instability is the so-called edge localized mode (ELM), which affects the periphery of the plasma leading to the expulsion of plasma filaments that impact the wall-protective elements (figure 6.32(A)). If not controlled, in some cases local instabilities can grow and eventually trigger global instabilities.

The basic physics processes behind such instabilities, how to avoid them, and approaches to control them are well established. However, challenges remain for ITER and future fusion reactors because of the specificities of self-heated plasmas and because of the high plasma pressures which are required to achieve effective fusion power production. The latter implies the need to maintain the plasmas in the stable equilibrium zone but near stability boundaries. The presence of energetic helium ions (and from external heating sources) in the plasma can cause specific instabilities (so-called Alfvén instabilities) [98] that can cause the expulsion of these energetic ions from the plasma and reduce plasma self-heating and fusion power production. These need to be avoided or reduced to a sufficiently low level that the impact on fusion production is small. Operation of fusion reactors near stability boundaries implies the need to quantitatively predict these boundaries and to develop the capabilities to control the parameters of the plasma by external actuators (external heating, magnetic fields, etc) to avoid overstepping them; this is particularly complex in self-heated fusion plasmas because of the feedback loops described previously. Finally, if uncontrolled instabilities develop, it is necessary to mitigate their consequences. A prototypical example is the tokamak disruptive instability, which is a global instability in which the plasma loses its thermal energy in timescales of milliseconds and its magnetic energy in timescales of tens of milliseconds, leading to large forces being exerted and large plasma fluxes being deposited on the tokamak wall and components.

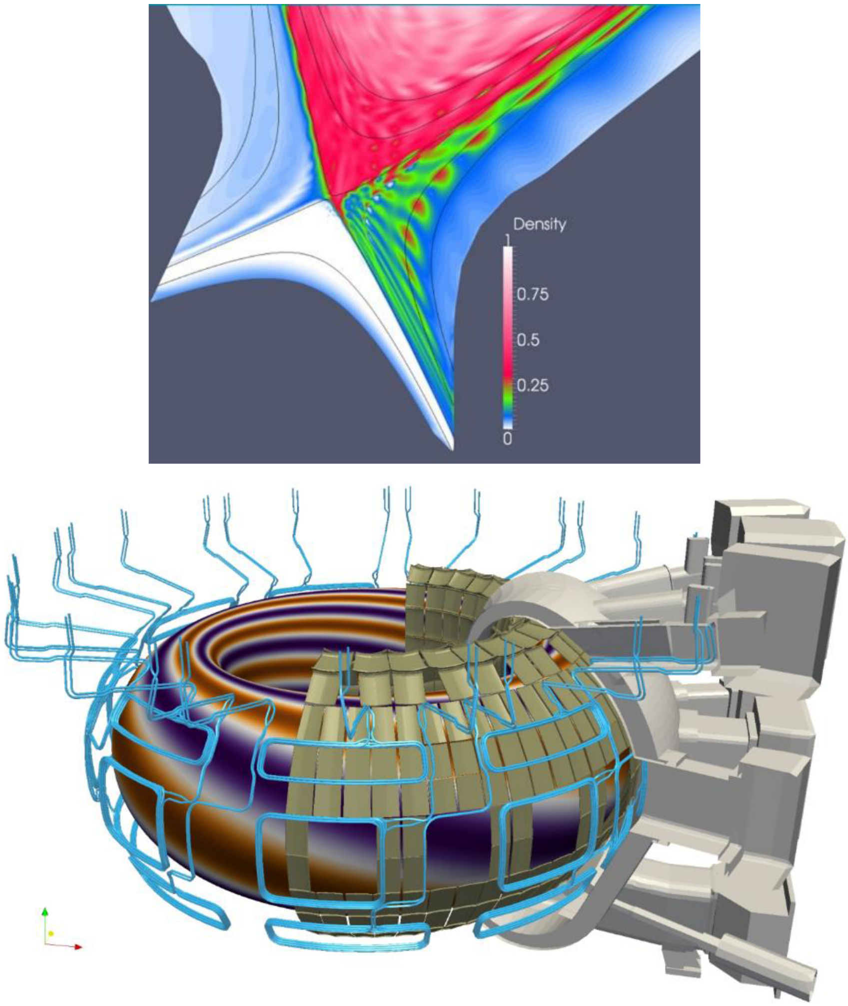


Figure 6.32. (A) Modeled plasma density contours showing the expulsion of plasma filaments from the peripheral plasma to the wall by ELMs in ITER. (Source [99].) (B) Set of 27 coils (in blue) to control ELMs in ITER. (Credit ITER Organization, courtesy of G Huijsmans.)

A major scientific objective of ITER is thus to demonstrate the avoidance and control of plasma instabilities in self-heated fusion plasmas and the mitigation of their consequences when these cannot be avoided. To achieve this goal, ITER is equipped with a set of versatile systems to create local currents in the plasma (by selectively acceleration of electrons with radiofrequency waves) to modify locally the magnetic fields at the plasma periphery (see figure 6.32(B)), and so on. Specifically for disruptions, ITER is equipped with a sophisticated system to reduce the heat fluxes and forces when disruptive instabilities cannot be avoided by the injection large amounts of hydrogen and neon in small solid pellets. This injection leads to a transient increase of the plasma density that allows the thermal and magnetic energy of the plasma to be lost as a flash of light emission (thus the need for neon) that is uniformly deposited on the tokamak walls rather than by plasma fluxes and electric currents on the tokamak wall and components.

Integration of fusion power producing plasmas with tokamak wall requirements
As discussed in the previous section the energy lost by the plasma by conduction and convection is ultimately deposited on the tokamak walls by plasma electrons and ions. Since both move along magnetic field lines, this causes concentrated heat fluxes in a very small fraction of the total wall area. For ITER and future fusion reactors these power fluxes are comparable to those at the Sun's surface (~60 MW m^{-2}), and there is no technological solution to evacuate such levels of heat for any significant length of time. For example, ITER's wall components are cooled by a high-pressure water flow and have demonstrated capabilities to extract heat fluxes up to 10 MW m^{-2} for long durations, but higher heat fluxes, when sustained over time, cause component degradation and reduction of lifetime. In addition, the impact of a high-temperature plasma causes erosion of the wall by energetic ion impact, similar to sandblasting of a surface, and this also limits its lifetime. Replacing wall components in ITER and fusion reactor is a complex operation that can take several months and thus reduces the availability of the reactor for experiments (ITER) or electricity production (DEMO and later fusion reactors). Therefore, wall component degradation and erosion need to be minimized.

The solution to this problem has been demonstrated in present experimental facilities by the injection of so-called impurities such as noble gases (e.g., neon) which transform the heat flux from the plasma into light, which deposits over a wide wall area, in a similar way to a fluorescent lamp. The intense light emission decreases the local heat fluxes on the wall and cools the peripheral plasma in contact with the tokamak wall down to temperatures of ~10 000 K rather than the several million kelvins that the plasma in contact with the material would have otherwise. This large decrease of plasma temperature implies that the ions impacting the wall do so at very low energy and erosion of the wall is thus negligible. On the other hand, if the impurities for heat flux dissipation enter the core plasma from the periphery where they are injected, they can cool and dilute the deuterium–tritium plasma and decrease fusion power production. To optimize heat flux dissipation versus impurity penetration into the core plasma, specific configurations of the peripheral magnetic field have been explored in present experimental devices (so-called divertor configurations) and have been adopted for ITER and future fusion reactors. The challenge remains for ITER to demonstrate that this approach can provide the required heat flux dissipation and sufficiently 'clean' plasmas for high Q fusion power operation. An important ingredient in this challenge concerns the understanding of the physics processes that govern the transport of plasma particles (including impurities) and heat at the plasma periphery. It is unclear whether the same processes that govern these phenomena in present experiments will be at work in ITER or whether other processes will dominate peripheral plasma transport, given the large densities, temperatures, and gradients expected at the ITER plasma edge.

Therefore, for ITER to achieve its goals it is necessary to demonstrate that a core plasma, with the high temperatures and density required for deuterium–tritium fusion to take place, can coexist with the peripheral plasma conditions required to have fluxes on the wall that can be handled by present technologies in terms of both heat fluxes and erosion. This requires a detailed understanding and control of the interaction between the plasma and the wall materials, the dynamics of the injected impurities, and so on. This is a prototypical example of the core–peripheral plasma integration issues that need to be demonstrated in ITER and future fusion reactors, but there are many others issues, such as the exhaust of the helium ash and the control of power fluxes due to ELMs. In general, such integration issues become more challenging for fusion reactors beyond ITER because of the higher fusion powers required for the demonstration of electricity production. In this regard, ITER is an essential first step on the way to the solution of such integration issues, but additional steps beyond ITER will be required to demonstrate the degree of integration required for future fusion reactors.

To conclude this subsection, we note that the challenges that we have discussed focus on plasma and plasma–material interaction physics, but these are not the only issues that need to be faced for fusion energy development. For example, it is necessary to understand the behaviour of materials to be used for the construction of fusion reactor vessels and in-vessel components under the high-energy neutron fluxes produced by deuterium–tritium plasmas and to demonstrate *in situ* tritium production (from lithium) to the level required for self-sufficiency in a reactor. ITER will provide a first insight into such challenges and an initial demonstration of tritium production, but the knowledge required for future fusion reactors will need to be gained through future dedicated experimental facilities (see [100, 101] for a description of these issues and challenges).

6.3.4.3 Conclusions

Fusion energy is a very attractive option to supply the energy needed by humankind in the next centuries. The fuels required are abundant, and the fusion reaction produces no long-lived radioactive residues. The process of research and development of fusion energy has been a long one that started in the middle of the 20th century and has now come to fruition. The next few decades will be crucial to demonstrate the scientific and technical viability of fusion as an energy source by integrating the acquired knowledge in physics optimization (e.g., energy and particle confinement, impurity control, power exhaust) and engineering optimization (e.g., plasma-facing components, tritium technologies, heating and current drive systems, plasma diagnostic systems). A key step in this demonstration will be the scientific exploitation of the ITER experimental reactor in which deuterium–tritium fusion plasmas, producing more energy by fusion reactions than that required to heat them, will be demonstrated for the first time. This key step will be accompanied by research in other specialized experimental facilities to address additional challenges (e.g., neutron resistant materials, advanced power exhaust schemes), which need to be addressed for the design and construction of a net electricity-producing fusion reactor.

The successful resolution of these integration challenges for fusion energy development requires further understanding and control of the physics processes that govern the behaviour of fusion plasmas. As an example, such detailed understanding is essential for maximizing power production in fusion reactors while ensuring an acceptable lifetime of the wall-protective elements whose power-handling capabilities are determined by materials properties and cooling technologies limits. This advance in physics understanding will be the focus of experimental research in the next decades for the new generation of experimental fusion devices that have recently come into operation or are presently under construction. Among them, ITER will provide unique contributions for self-heated deuterium–tritium reactor-scale plasmas, which are essential for the follow-up step of a net electricity-producing fusion reactor such as DEMO. This experimental research will be accompanied by theory and modeling developments, taking advantage of the capabilities provided by advanced supercomputers, which are necessary to understand the details of the physics processes at play in experiments and for their extrapolation.

6.3.5 Fission energy: general overview

Marco Ripani [1]
[1] National Institute for Nuclear Physics, Genova, Italy

There are three basic physical phenomena behind nuclear energy from fission. First of all, some specific heavy nuclei, called *fissile*, can easily split into two lighter fragments when they absorb a neutron[1]. Second, when such nuclei split, more neutrons with relatively high energy of motion are emitted, which can in turn produce other fissions, in a so-called *chain reaction*. Finally, thanks to Einstein's mass–energy equivalence, the total mass of the two fragments and the few neutrons emitted is less than the initial mass of the fissile nucleus; the missing mass has been transformed to energy. Each fission reaction releases about 3.2×10^{-11} J. By comparison, a chemical reaction releases energy of the order of a few 10^{-19} J.

In nature, the element uranium is found in two different forms, or isotopes[2], namely, U-235 and U-238, in the proportion of 0.7% for the first and 99.3% for the second. U-235 is fissile and can undergo fission with high probability when exposed to slow neutrons. Nuclei such as U-238 are called *fissionable*, that is, they can give rise to a fission reaction, but only for fast neutrons above a certain energy of motion. Nuclei such as U-238 are also called *fertile*, because when hit by a neutron, they start a series of nuclear reactions that lead to the formation of new elements. In the case of U-238, irradiation with neutrons leads to the formation of plutonium isotopes, among which the most relevant is Pu-239, which is fissile. Therefore, starting from a mixture of U-235 and U-238, not only fission of U-235 occurs, but also production, or *breeding* of new fuel, in terms of the mixture of plutonium nuclei containing Pu-239. Most reactors using mixed U-235/U-238 fuel typically need 3%–5% U-235, so a procedure called uranium *enrichment* is needed to fabricate reactor fuel. Other reactors work by using natural uranium without enrichment. It is worth mentioning that thorium, an abundant element in nature, can in principle be used to produce (or breed) a different type of fuel. Indeed, in neutron absorption by thorium the final product is U-233, yet another fissile nucleus. Use of thorium as a breeder of U-233 is studied internationally as a possible way to increase the available stock of fissile fuel.

The whole process of mining, extraction of the uranium mineral from the ore, purification, fuel fabrication through enrichment, irradiation, and final storage or disposal is called the *fuel cycle*. In the *open* or *once-through* fuel cycle the spent fuel is put in temporary storage waiting for final disposal. In the *recycling* (sometimes called *closed*) fuel cycle the spent fuel is reprocessed to extract the plutonium, which is then used to fabricate new fuel, typically in the form of mixed oxide of uranium and plutonium.

[1] The neutron is one of the two components of atomic nuclei (protons, which are positively charged, and neutrons, which have no electric charge).
[2] Isotopes are nuclei of the same chemical element that contain the same number of protons but different numbers of neutrons.

The fragments emerging from fission are typically radioactive, so the fuel extracted from the reactor, *irradiated* or *spent* fuel, is a highly radioactive material that needs to be protected and shielded. While most of the radioactive fragments have relatively short half-lives[3], the longest being of the order of a few tens of years, a small group of fragments has very long half-lives of the order of 10^5 years, the so-called long lived fission products. The aforementioned plutonium production process is one example of a process leading to the appearance of *transuranics* in the fuel, that is, chemical elements beyond uranium that do not exist in nature. Many transuranics have long half-lives, from a few hundred years to a few hundred thousand years (e.g. Pu-239 has a half-life of 24 000 years). Reprocessing the irradiated fuel gives the possibility of recycling the Pu to fabricate new fuel, but whenever reprocessing is not applied or is stopped, the resulting spent fuel is considered to be *nuclear waste* or *radioactive waste* which needs to be properly stored and eventually disposed of. Indeed, such radioactive nuclei can be dangerous for the environment and for human health (due to direct exposure, ingestion, or inhalation). Additional radioactive waste is produced in the reactor by neutrons interacting with the surrounding materials, producing *activated* materials with varying degrees of radioactivity. The International Atomic Energy Agency has come up with a classification scheme that is used as a basis by each country to define categories of waste [102].

In terms of fuel consumption, in a 1-GWe reactor[4] at 80% load factor (the actual operational time) the annual consumption of fissile material is about 900 kg: in volume of pure metallic heavy elements, this would be a cube with sides of about 35 cm. If we consider an actual oxide fuel enriched in U-235 to be 3.5%, the annual fuel consumption will be of the order of 30 tonnes. A comparison between the annual consumption of different fuels is shown in table 6.1[5].

All nuclear reactors today (so-called Generation II, III, and III+) are based on fuels containing U-235 and/or Pu-239. With a few exceptions (to be discussed later in this section), they contain light materials such as water, *heavy water*[6] or graphite to slow down the fast fission neutrons and increase the fission probability[7]. Water or

Table 6.1. Comparison between different annual fuel consumption rates, for a 1-GWe plant at 80% load factor.

Nuclear fuel (tonnes)	Natural gas (cubic meters)	Coal (tonnes)
30	1.4 billion	2.1 million

[3] Radioactivity declines with time following an exponential law. The time after which the amount of radioactivity has halved is called the half-life.
[4] A reactor producing about 3 GWth of thermal power converted to about 1 GW of electric power at 33% conversion efficiency.
[5] Assuming a 50% conversion efficiency for gas and 40% for coal.
[6] In heavy water, instead of the ordinary H_2O molecule, hydrogen is replaced by its heavier partner deuterium.
[7] These materials are called *moderators.*

gas is used to cool the fuel and transport the heat produced by the fission reactions to the equipment, producing electricity. In *fast reactors*, the materials crossed by the neutrons are chosen to keep the neutrons fast and energetic. This can be accomplished if the core contains mostly heavy materials. Therefore, for core cooling, the choice will be a liquid metal such as sodium, lead, or a lead–bismuth mixture; alternatively, the coolant should be a low-density material such as helium gas. On the other hand, the fission probability is smaller for fast neutrons, which means that a much higher percentage of fissile material, of the order of 20%, is needed to sustain the chain reaction. Only two examples of fast reactors connected to the electricity grid exist today (to be discussed later in the section), while many others are prototypes or research reactors, and a few more are planned for the future.

6.3.5.1 Worldwide figures

Worldwide electricity consumption increased annually by 2.9% on average in 2007–20 [103], regardless of progress in transmission or utilization efficiency, while a slight decrease occurred in 2020 due to the COVID-19 pandemic. Nuclear energy from fission continues to have an important share in the electricity mix. After recovering from the significant post-Fukushima drop due to several shutdowns and stress tests, production returned to pre-Fukushima levels, apart from COVID-affected 2020, with 2% more electricity supplied in 2019 with respect to the average of 2007–10 (see figure 6.33). However, while total electricity production increased by 31% in the same years, the worldwide percentage of nuclear electricity went from about 14% to around 10% [103, 104] (see figure 6.34). As of July 2021, there were 443 reactors in operation in the world, for a total net installed capacity of about 393 GWe, and 51 under construction, for a total capacity of about 53.9 GWe [105].

6.3.5.2 Cost of electricity

One of the parameters used to compare the economics of different electricity technologies, independent of their very different characteristics, is the levelized cost of electricity (LCOE). It is calculated by dividing costs of a plant over its whole

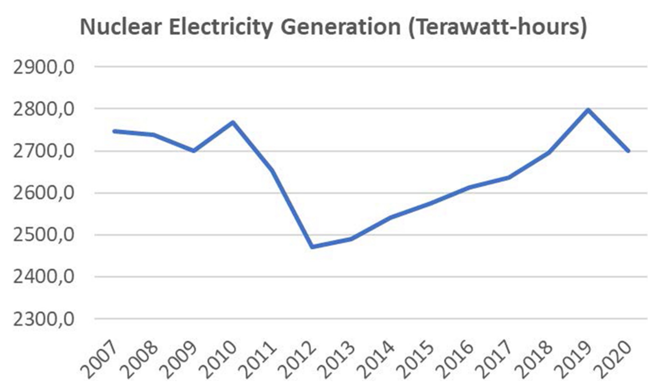


Figure 6.33. World nuclear electricity generation from 2007 to 2020. Data from [103].

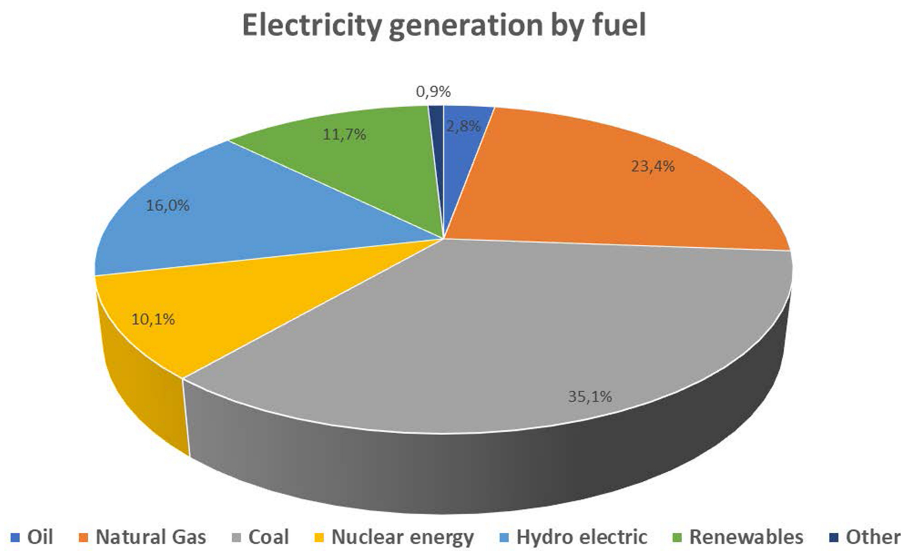


Figure 6.34. World electricity generation by fuel (2020). Data from [103].

lifetime by energy production and is typically given in USD/MWh. In the IEA/NEA/OECD study of 2020 [106], at a 7% discount rate, nuclear energy's LCOE appears to be definitely lower than that of coal and combined cycle gas turbines with carbon capture and storage, as well as lower than that of most renewables, with the exception of utility-scale photovoltaic, onshore wind, and hydropower. When considering lifetime extensions of existing nuclear plants, the nuclear LCOE is definitely lower than any other source, be it coal or renewables, with only a marginal superposition with utility-scale photovoltaics.

A similar study was performed by the Intergovernmental Panel on Climate Change (IPCC) [107] that assumed a 10% discount rate scenario, two scenarios for the number of operational hours of the plants, and two scenarios for the cost of carbon emissions for the CO_2 emitting technologies. This study, too, shows that nuclear energy features an interesting LCOE, especially when carbon emission costs are factored in, and is cheaper than many renewable or frontier technologies. Therefore, the nuclear LCOE appears to be competitive with that of other sources. This is particularly true in low discount rate scenarios, given that nuclear energy is a capital-intensive energy source, requiring large investments at the construction stage.

6.3.5.3 Carbon emissions

Although nuclear power plants do not emit CO_2 (nor other greenhouse gases or fine particulates) during electricity production, CO_2 can be emitted along the whole plant lifetime, from fuel mining and extraction to final decommissioning. A detailed study comparing emissions among electricity production sectors in terms of CO_2 per unit of produced energy has been performed by the IPCC [107]. Their report clearly shows that the small amount of CO_2 produced as a result of the whole life cycle is much lower than that from fossil fuel–fired power plants, is comparable with that of many renewables, and is somewhat lower than that of solar photovoltaic. A similar conclusion is found in [108] for the case of the twice-through cycle adopted in France, in which the nuclear spent fuel undergoes one cycle of reprocessing to fabricate new fuel from uranium and plutonium, with 5 g of CO_2 equivalent production per kilowatt-hour of nuclear electricity produced, the smallest value in a group including coal, oil and gas, photovoltaic, hydropower, and wind.

6.3.5.4 Resources

World annual uranium consumption amounted to 62 825 tU (tonnes of uranium) in 2017, and assuming it to be constant, identified recoverable resources should suffice for over 130 years [109]. It is estimated that additional undiscovered and unconventional resources could suffice for well over 400 years, but this would require significant efforts in securing all resources for an effective use. Moving to advanced technology reactors and recycling fuel could increase the long-term availability of nuclear energy from hundreds to thousands of years. Thorium, which is more abundant than uranium in the Earth's crust, is another potential source of alternative fuel if used to breed U-233, a fissile nucleus that does not exist in nature. Even when relying only on identified resources, assuming an increase in nuclear energy production as in the IEA NZE scenario and considering only current reactor technologies, there should be enough uranium supply for the rest of the century. This implies that nuclear energy from fission can have the important role of a bridging technology until nuclear fusion becomes commercially available. However, research on fuel recycling and innovative reactor technologies should continue to be pursued to have the possibility to extend the availability of nuclear power from fission well beyond the end of the century. This is also important because both fuel recycling and the introduction of innovative reactors, in particular fast reactors, would help to reduce the amount of radioactive waste to be disposed of.

6.3.5.5 Accidents and safety

A lot of public concern has been raised by nuclear accidents, most notably the ones at Chernobyl in 1986 and Fukushima in 2011. Obviously, the use of nuclear power must be appropriately regulated and managed with maximum care, given the potential impact of accidents entailing release of radioactive substances into the environment. Even though no industrial activity (and no human activity in general) can be made totally risk-free (indeed, several accidents have happened in the fossil fuel sector, including oil spills in the sea, fuel truck and train carriage explosions, and fires), the goal of nuclear plant designers is to incorporate as many safety measures as possible to prevent any anomaly from escalating to a significant nuclear accident, following the principle of defence in depth [110]. Safety is an aspect of utmost importance in nuclear power production. The core of a nuclear reactor produces radiation both while in operation and after shutdown, because of the high content of radioactive materials in the fuel and because of the neutron irradiation of structural materials, which produces radioactive nuclei starting from stable ones. Lessons learned from past accidents have been implemented in the design of plants, with particular regard to engineered safety measures. Attention to safety is present from the design stage all the way through to operation to shutdown and decommissioning of the plants, and particular care is devoted to the analysis of possible accidental scenarios. A fundamental principle is the *defence in depth* [110], which means that a series of barriers, starting from the solid fuel form itself up to the reactor containment building, are put in place to contain radioactivity as much as possible at all times and in all instances. Moreover, research specifically focused on the safety of nuclear plants is being conducted worldwide to pin down and address residual weaknesses, also taking advantage of lessons learned from the past plant operational history. Obviously, in this respect, the role of independent regulatory authorities is crucial in overseeing all stages of a plant lifetime, from design to operation to decommissioning and dismantling, and corresponding radioactive waste management. Clearly, scientific and technical aspects are the main practical objectives of a safety assessment. However, human and organizational factors are also deemed to be of high importance, as, for example, human errors may still have an impact even in a system which is designed and operated according to the state of the art. Therefore, what is called culture for safety means that technical achievements and established engineering standards are a must, but a continuous process of self- and reciprocal assessment on human and organizational aspects is also necessary, involving all actors, from industry to regulators, including the stakeholders for whose benefit nuclear energy is ultimately utilized [111].

6.3.5.6 Radioactive waste

The decay time of fission products is the reason why disposal of spent fuel and waste requires very long term storage, for which one possible solution envisaged is storage deep underground in the so-called geological repositories. According to the IAEA categorization [102], depending on quantity and decay time, nuclear waste is classified into Ex-emptExempt waste (EW), Very short lived waste (VSLW), Very low level waste (VLLW), Low level waste (LLW), Intermediate level waste (ILW), and High level waste (HLW). VLLW is suitable for disposal in near surface facilities of the landfill type with limited regulatory control. LLW requires robust isolation and containment for periods of up to a few hundred years and is disposed of in engineered near-surface facilities. ILW requires a greater degree of containment and isolation, which can be provided by disposal at greater depths, of the order of tens of metres to a few hundred metres. For HLW, disposal in deep stable geological formations, usually several hundred metres or more below the surface, is the generally recognized option for disposal. In the case of HLW, the internal heat generated by radioactive decay may have to be taken into account. Near-surface disposal facilities are safely in operation since many years in several countries around the world [112]. The IAEA estimated that, since the start of the nuclear power era in 1954 to the end of 2013, a total of about 370 000 t HM (tonnes of heavy metal) of spent fuel was discharged from all nuclear power plants worldwide (excluding India and Pakistan), of which about one third was reprocessed to produce a second round of energy production with recycled fuel [113]. This total amount of discharged spent fuel would correspond to a few meters over the surface of a standard soccer field. International recommendations (e.g. [114]) have guided national policies, strategies and programs for the management of spent fuel and radioactive waste. Surface or near-surface facilities are a reality in several countries and demonstrated that safe disposal of the corresponding classes of waste can be easily realized. Continuing research on the optimal choices of sites and the prediction of the time evolution of the waste packages underground is performed at international level. In particular, Euratom has promoted the advancement of research on several aspects of decommissioning technologies (e.g. SHARE Euratom project [115]), as well as on aspects and issues arising in interim storage, pre-disposal and final disposal, where in the latter point the scientific aspects of deep geological disposal are of special interest. Alternatively to geological disposal, it has been proposed to burn at least part of the nuclear waste with accelerator-driven systems, based on a subcritical reactor core [116]. A subcritical core is one where the fission chain reaction does not occur spontaneously in a steady state but rather requires an external neutron source to work, such as that produced by an accelerated beam of protons impinging on a thick target. Thanks to the subcriticality, there would be the possibility to mix a certain quantity of minor actinides (MA) into the fuel, which make it possible to incinerate them by fission. This is called transmutation of the MA into fission products, which can reduce the radiotoxicity of the final materials in the fuel in the long term, such that the confinement time would be reduced by orders of magnitude, from hundreds of thousands of year to a few hundred years. In order to set up such an incineration cycle, both partitioning and transmutation are needed. In the partitioning stage, plutonium and MA are separated from the spent fuel; then in the transmutation stage they are irradiated in a special core to be incinerated [117]. Supporting a robust research program is necessary because radioactive waste management programs present special challenges in their planning and execution. Indeed, besides involving science and technology, they also involve programming, regulatory, public participation, and stakeholder involvement aspects [118]. While engineering aspects are generally well understood, the necessary scientific topics of site characterisation, process modeling, safety assessment, and so on can evolve over time and require the capability of adapting the program planning to new findings [118].

6.3.5.7 Nuclear power in future scenarios

There are several projections about the future development of electricity and energy supply sources. Nowadays, several scenarios consider how to develop a low-carbon economy [119]. Typically, nuclear energy plays a role in all projections, even when a general trend towards phasing out nuclear power is conservatively assumed. Indeed, nuclear power plants are capable of delivering a large quantity of electric power (typically a power of the order of 1 GWe with high availability), providing a solid contribution to baseload electricity production. Like hydropower, nuclear power requires much less material resources for the same electricity output, due to high power density and capacity factor. As a result, nuclear power plants have relatively low emission levels per unit of energy produced [107, 108]. Nuclear plants provide dispatchable electricity and can be regulated according to the system needs, although rapid load following is not yet a common feature of these types of plants. Moreover, they have the potential to offer grid services needed to ensure continuous balance between load and generation [119].

The IPCC report on limiting global warming within 1.5 °C [120] discusses four illustrative pathways by which different measures are adopted across energy production, industry, and agriculture to mitigate the temperature increase on the planet. In all four pathways, nuclear energy is expected to increase in 2030 with respect to 2010 by amounts that vary between 60% and 100%. Variations from 2010 to 2050 are even higher, between 100% and 500%. Even though there is a large variation among the many different models considered, 50% of the models indicate an increase in nuclear energy from 2010 to 2030 between 44% and 102%[8], while 50% of the models indicate an increase in nuclear energy from 2010 to 2050 between 91% and 190%.

Among the most recent studies is the IEA Roadmap to Net Zero [121]. This study considers three scenarios: the Stated Policies Scenario (STEPS), based on well-established policies and up-to-date information about energy, climate plans, and related policies based on explicit laws or regulatory actions, where the latter also come from commitments following the Paris Agreement and can cover economic measures as well; the Announced Pledges Case (APC) scenario, based on timely and full realization of all national net-zero emissions pledges; finally, the Net-Zero Emissions by 2050 scenario (NZE) discusses what is actually needed, and when, to achieve net-zero CO_2 emissions by 2050 worldwide in energy production and industry. In both STEPS and APC, nuclear power's role in the total energy supply will increase from now to 2050. In STEPS, nuclear energy is projected to grow by 15% between 2020 and 2030 and still more until 2050, while in APC the nuclear growth is projected to be around 25% by 2030, so nuclear energy would maintain its 10% share of electricity production and would maintain a similar trend until 2050. In NZE, the role of nuclear power is essential, and its contribution to the total energy supply is projected to increase by nearly 100% between 2020 and 2050.

6.3.5.8 Obsolescence of the reactor fleet and economics

Most reactors in operation today were built between the 1970s and the 1980s [122], and the typical lifetime of a plant ranges between 40 and 60 years. As a consequence of that and of the fact that some countries decided to phase out nuclear power, many plants are already being decommissioned or will be in the next couple of decades worldwide. On the other hand, in recent US and Western Europe experience, construction costs of nuclear power capacity have increased from about 2000 USD/KWh in 2005 to between about 7000 and 8000 USD/KWh [122]. Moreover, for

[8] This is the interquartile range of the models, meaning that models predicting variations below 44%—even negative ones where nuclear energy decreases—represent 25% of the total, while another 25% predict variations above 102%.

example, in the US, nuclear power is experiencing a strong market pressure due to low wholesale electricity prices, driven by low gas prices and rising renewable power capacity (see, for instance, [119]), yet another economical aspect that is affecting the ability of the utilities to remain profitable and stay on the market producing nuclear power.

6.3.5.9 Innovative reactors

In the effort to improve the safety, security, and efficiency of nuclear plants, new concepts of reactors have been developed with goals including the minimization of the production of MAs, a better and more efficient use of the fuel, a better thermodynamic efficiency, the possible production of hydrogen at high temperatures and finally improved safety features to minimize the risk of accidents. All these various concepts are considered within the so-called Generation IV reactors, which are the subject of an international initiative [123] and are one of the pillars of the European Sustainable Nuclear Energy Technology Platform [124]. In particular, fast reactors use liquid metals or gas as coolants, resulting in more energetic neutrons traveling within the reactor core. In such reactors, it becomes possible to partly burn not only fissile elements such as U-235 and Pu-239, but also fissionable elements such as U-238 and some transuranics, for which fission occurs only above a certain minimal neutron energy. Typically, the fuel has to be richer in U-235 or Pu-239 content, but on the other hand, to some extent U-238 becomes a fuel as well, which means that uranium resources are used more efficiently and will last longer. Another important feature of fast reactors is that the ratio between fission and production of transuranics is higher, so relatively lesser amounts of transuranics are formed. At the same time, thanks to the more energetic neutrons, in a fast reactor the transuranics that are produced can be partly incinerated in the reactor itself by fission, resulting in less long-lived radioactive waste. Besides several prototypes based on different technologies around the world, two sodium-cooled commercial reactors have been in operation for several years in Russia [125].

Recently, there has been growing interest in the concept of small modular reactors (SMR), delivering up to about 300 MW of electric power, with several ongoing projects around the world [126, 127]. Besides advantages in terms of improved safety features and ease of deployment due to high standardization, SMR are getting serious attention for district heating, where a major impact on CO_2 emissions can be expected.

6.4 Towards green cities: the role of transport electrification

Natalio Mingo[1], Gérard Gebel[1], Philippe Azais[2], Thierry Priem[2], Tuan Quoc Tran[3], Didier Jamet[1], Florence Lefebvre-Joud[1] and Simon Perraud[1]
[1]Université Grenoble-Alpes, CEA, Liten, Grenoble, France
[2]Université Grenoble-Alpes, CEA, DPE, Grenoble, France
[3]Université Grenoble-Alpes, CEA, Liten, INES campus, Le Bourget du Lac, France

6.4.1 Transport, climate change, and air pollution

Global warming has reached a point where there is widespread concern about the large impacts that are being observed on ecosystems, human health, food security, and water supplies [128]. CO_2 constitutes the main source of anthropogenic global warming, representing around 70% of greenhouse gas emissions [129]. Global CO_2 emissions are rising steadily [130], being about 60% higher in 2018 (33.5 Gt[CO_2]) than in 1990 (20.5 Gt[CO_2]) when the first IPCC report was edited. The transport sector, which relies heavily on fossil fuels (mainly oil), accounts for one fourth of the global CO_2 emissions.

Apart from CO_2 emissions, transport has also a major impact on air pollution and hence on human health [131]. The combustion of fossil fuels for transport applications leads to the emission of a range of air pollutants, such as particulate matter (PM), nitrogen oxides (NO_x), and carbon monoxide (CO), which have a significant impact on human health, reducing life expectancy and increasing medical costs. The European Environment Agency estimates that in 2018, long-term exposure to particulate matter with a diameter of 2.5 μm or less ($PM_{2.5}$) led to around 379 000 premature deaths in the EU-28 [132].

When comparing various cities in countries having similar GDPs [133], the example of Hong Kong stands out. The city has a rate of energy consumption of 80 kWh/person-day, significantly smaller than that of cities in France (150 kWh/person-day) or in the US (>200 kWh/person-day). Being densely populated, Hong Kong has an efficient public transportation, compact housing, and almost nonexistent private car ownership; these are presumably part of the explanation. This suggests that there is a lot that infrastructure and lifestyle, in particular transport, can help with when it comes to achieving greener cities.

6.4.2 How to reduce urban transport emissions?

In the EU, significant efforts have led to energy efficiency improvements in the transport sector. For example, in the case of road transport, the average CO_2 emissions from new passenger cars [134], measured in g[CO_2]/km, decreased by 30% between 2000 and 2018. However, transport demand continues to grow, and the sector remains the only one that has not been able to reduce its CO_2 emissions. In the EU, transport emissions have increased by around 20% compared to 1990 levels, while the EU total emissions have decreased by around 20% in the same period.

In order to lower the CO_2 and air pollutant emissions of urban transport sector, several routes can be considered in parallel, including the following:

1. Replacement of motorized transport by nonmotorized transport modes, such as bicycles (generating no emissions).
2. Replacement of individual transport by public transport modes, such as trains, trams, subways, and buses (generating much lower emissions, on a passenger and kilometer basis, than passenger cars).
3. Development of modular mobility with multimodal transportation networks, mixing public and individual transport modes with digital management of fleets and fluxes.
4. Replacement of internal combustion engine vehicles by electric vehicles.

Here, we focus on the last of these routes, that is, the electrification of transport using batteries or hydrogen fuel cells. The main advantage of electric vehicles (EV), such as battery electric vehicles (BEV) or hydrogen fuel cell electric vehicles (FCEV), compared to vehicles with internal combustion engine (ICE), is that EVs have much lower local emissions (since there is no combustion of fossil fuel), and therefore they are very good solutions to reduce air pollution, notably in urban areas. Furthermore, if the electricity used to recharge the battery of the BEV (or to synthesize the hydrogen by water electrolysis for the FCEV) is made from low-carbon primary sources (nuclear or renewable energy), then the electric vehicle has a low life-cycle carbon footprint compared to fossil fuel–powered vehicles, hence contributing to significantly reducing CO_2 emissions and fighting against climate change.

6.4.3 Transport electrification: state of the art

6.4.3.1 Electric vehicles and their powertrains

The electrification of vehicles has a long history. At the beginning of 20th century, there were more EVs in New York than ICE vehicles. However, the rapid progress of the latter, an attractive range of vehicles, and the development of the gas station network quickly overtook EVs. Since the 1990s, the emergence of efficient electrochemical storage solutions, the threat of fuel dependency, and increasing concerns about pollution have brought renewed interest in partial or complete electrification of vehicles' powertrain.

In the case of hybrid electric vehicles (HEV) (figures 6.35 and 6.36), the high efficiency of the electric motor (>90%) and its very high torque even at low speed are advantageously exploited to reduce emissions in the torque/engine speed range where the combustion engine has poor efficiency. Even if the only fuel remains fossil, it is used more efficiently. HEV are currently undergoing strong commercial development and should rapidly replace the NiMH battery-based 'full hybrid' solutions that have been widely marketed since 1997 (especially by Toyota).

A more complex solution, the plug-in hybrid electric vehicle (PHEV), generally a parallel hybrid type, aims at coupling two solutions requiring two energy sources: electricity from the grid and a fossil fuel. This solution makes it possible to travel a significant number of kilometers in all-electric mode (typically 50 km), which

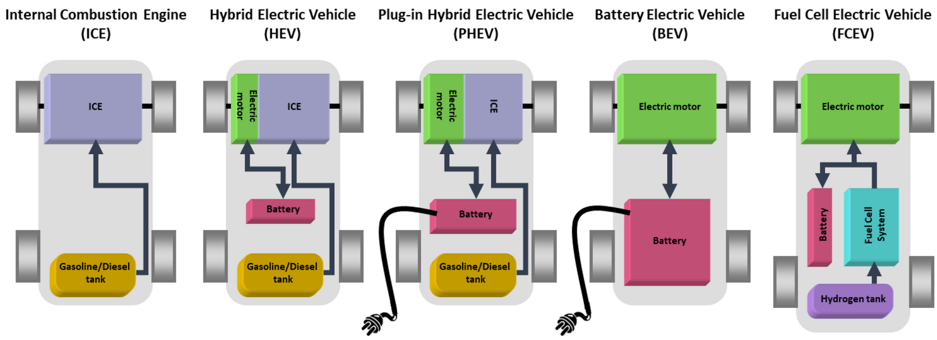


Figure 6.35. Simplified layouts of different powertrain configurations. (This figure is from the CEA HPC website (http://www-hpc.cea.fr). All reproduction rights are reserved.)

E-powertrain type		Micro	Mild / HEV	Full	PHEV	BEV
Functions	Start Lighting Ignition	X	X	X	X	
	Start & Stop	X	X	X	X	
	Boardnet	X	X	X	X	X
	Brake energy regeneration		X	X	X	X
	Boost		X	X	X	X
	Motor downsizing		X	X	X	
	ZEV		(a few 100m)	a few km	40 to 120km	> 150km
Main characteristics	System voltage	12V	48V	<300V	200 to 400V	300 to 800V
	System Power	< 5kW	10 to 30kW	40 to 165kW	> 30 kW	> 50 kW
	Usable energy/cycle	< 5Wh	0.3 to 1.3kWh	< 400Wh	5 to 18kWh	> 20kWh
	Installed energy	<2kWh	1 to 3kWh	< 2kWh	6 to 20 kWh	> 22kWh
	CO_2 decrease ratio	<5%	< 15%	< 20%	> 35%	100%
	Electrochemical Energy Storage System	EDLC, Pb Acid and / or Li-ion	Li-ion	NiMH	Li-ion	Li-ion

Figure 6.36. Main characteristics and functions of electrified powertrain types in passenger vehicles. Micro: micro-hybrid vehicle; Mild/HEV: Mild-hybrid vehicle/hybrid electric vehicle; Full: Full hybrid vehicle; PHEV: Plug-in electric vehicle; BEV: Battery electric vehicle. ZEV: Zero emission vehicle range. (This figure is from the CEA HPC website (http://www-hpc.cea.fr). All reproduction rights are reserved.)

generally covers urban daily use. For longer trips, long-distance driving is possible using the ICE powertrain.

Finally, it is possible to electrify the powertrain completely, as in BEVs, totally getting rid of fossil fuels on board the vehicle. The major drawbacks of this solution are the charging time and the considerable mass of the Li-ion battery that is required for long journeys. Typical energy consumption of such a BEV is 10 kWh/100 km/ton in a Worldwide Harmonized Light Vehicles Test Procedure (WLTP) cycle [135].

In the case of FCEVs, a fuel cell is used to produce electricity for powering the electric motor. Hydrogen, which feeds the fuel cell, is stored in a high-pressure tank (usually 350 bar) containing a few kilograms of hydrogen[9]. In this configuration, the range of the FCEV is directly correlated to the size of the hydrogen tank. A battery is usually coupled to the fuel cell to manage power demands. This solution is therefore very close to a full-hybrid solution.

6.4.3.2 Brief introduction to batteries and fuel cells

In an electrochemical generator, an exothermal chemical reaction is converted into two separated electrochemical reactions involving the same ion shuttle and generating electrons.

In a Li-ion battery the ion shuttle is Li^+. A Li-ion battery cell consists of two electrodes (a cathode, that is, positive electrode, and an anode, that is, negative electrode) separated by a porous separator (to avoid short-circuiting between the electrodes) impregnated with an electrolyte (one or more organic solvents with a lithium salt dissolved in them). The electrolyte is the Li ion carrier between the two electrodes. The name 'lithium-ion' comes from the fact that Li ions are transported between the host electrodes duringthe charge and discharge cycles (figure 6.37).

The nominal voltage of a Li-ion battery cell is generally around 3.7 V. Therefore, it is necessary to combine several cells in series to increase the battery voltage and in

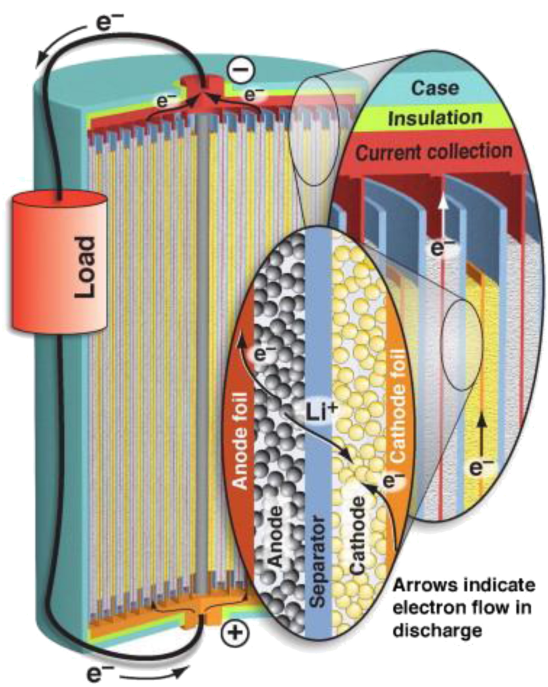


Figure 6.37. Principles of a lithium-ion battery cell. Reprinted from [136], copyright (2010), with permission from Elsevier.

[9] Approximately 1 to 1.5 kg of hydrogen are needed to travel 100 km.

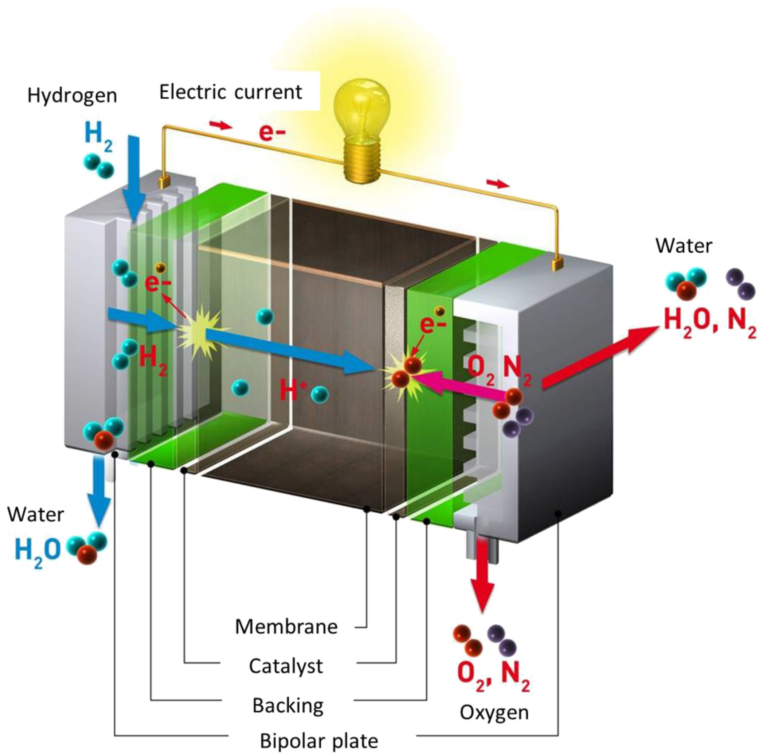


Figure 6.38. Principles of a PEMFC.

parallel to increase its capacity. In order to operate the battery in a safe and durable way, additional elements for thermal management, mechanical strength, safety devices, sensors, electronics, and so on are included in the battery pack.

In a proton-exchange membrane fuel cell (PEMFC), the ion shuttle is a proton H^+ (figure 6.38). Electricity is generated by separating the exothermal chemical reaction between hydrogen and oxygen into hydrogen oxidation at the anode and oxygen reduction at the cathode. Protons, produced on platinum-based catalyst at the anode, migrate through a proton exchange membrane (generally based on NAFION™ based polymer) to react with oxygen at the cathode and generate water vapor. Nominal voltage of a PEMFC individual cell ranges around 0.6 V; it is therefore mandatory to stack several cells to get required voltage, while current is directly adjusted with cells total active surface.

6.4.3.3 Current status of electric vehicles deployment

The electrification of passenger vehicles has been growing rapidly over the last 10 years. However, if the impact of transport electrification in urban areas is immediate, its global impact for reducing greenhouse gases is directly conditioned by the carbon footprint of the electricity that is used to recharge batteries.

At the end of 2020, the global vehicle fleet consisted of more than 1.2 billion vehicles, of which about 10.1 million were plug-in EVs (BEVs and PHEVs), with about 40% of the vehicles located in China. In the same year, BEV and PHEV sales reached 3.24 million vehicles, or 4.2% of new vehicle sales. The market shares of BEVs and PHEVs vary greatly from country to country. This is mainly due to very different financial incentive strategies, with some countries making a stronger choice for BEVs (France, South Korea, China) while others are pushing electrification in the broadest sense (PHEVs and BEVs), thus directing buyers towards PHEVs, as the latter are more suitable for all uses (UK, Sweden). The European car fleet consisted of 275 million passenger cars (EU28 + EFTA), of which about 1.9 million were EVs (BEVs and PHEVs). Of the 13.7 million new passenger cars sold in 2020, nearly 1.4 million were EVs (BEVs and PHEVs), or 10.2% of sales. It is worth noting the very high penetration of BEVs in Norway, representing 55.9% of new vehicle sales in 2019, compared with 5.3% of sales in the Netherlands, 2.8% in France, and 2.0% in Germany [137].

Commercial low-carbon solutions are available today for both light-duty and heavy-duty vehicles. However, while massive electrification of transport is becoming a reality, it does not solve the problems inherent in current transport modes such as the excessive number of vehicles (expected to remain so for a long time) and the limited energy efficiency in heavy traffic.

Public transport on dedicated tracks, such as tramways, trains, or subways, allows solutions which are largely decarbonised, as they are easily connected to electrical networks. Public transport solutions on open roads, such as buses, coaches, or trolleys, with on-board stored energy are already widely deployed (electric and/or fuel cell buses and coaches), and many demonstrators have been created in the last 10 years. At the end of 2019, there were about 500 000 electric buses (battery electric buses and plug-in hybrid electric bus) in operation worldwide. In 2018, more than 92 000 new electric buses were registered [138], mostly in China [139], compared to 104 000 in 2017. Plug-in hybrid electric buses account for 98% of new electric bus registrations. Outside China, about 900 electric buses were registered in 2018, mainly in Europe. China accounts for 98% of the global electric bus market. The city of Shenzhen had already completed the full electrification of its approximately 16 000 buses in 2017. European cities have only about 4000 electric buses altogether, and US cities have fewer than 400.

The World Resource Institute analysed the efforts of municipalities at different stages of electric bus adoption [140]. The main barriers to deployment were identified together with solution to overcome them:

1. Cities must first upgrade their electrical grid and develop a charging infrastructure. One of the difficulties in Shenzhen was the time it took to build a charging infrastructure capable of powering over 16 000 electric buses. For a range of about 200 km a bus needs to charge 250 kWh. In total, the fleet consumes over 4000 MWh per day.
2. The cost for acquiring a fleet of electric buses is often cited as the main obstacle. The price of a new electric bus ranges from US$300 000 to US$900 000 per unit. In the US, the average cost of an electric bus is US$750 000, compared to US$435 000 for a conventional diesel bus.

FCEV deployment appears to be significantly slower: About 4650 hydrogen buses were in circulation in the world by the end of 2020 [141], mostly in China with almost 4430 buses, in Europe with about 150 buses, and a little in the US with some 60 buses. This situation can be explained by several factors. BEV prices have decreased faster than expected, thanks to massive production. At the same time, numerous BEVs offer ranges of hundreds of kilometers (more than 600 km per charge for Tesla Model 3 or Xpeng P7). Finally, the deployment of fast charging stations (125 kW or more) allows an additional recharge of more than 200 km in less than half an hour. The differentiating factor in favor of FCEVs is their capability of withstanding very intensive daily use or daily long travels. These conditions can justify the deployment of expensive infrastructures (hydrogen pump, on-site hydrogen storage, etc).

6.4.3.4 Management of EV charging in smart cities

If not managed properly, the connection of a large number of EVs to the electrical grid can cause several technical problems for power systems, such as increasing peak demand and the risk of congestion and overload. However, EVs may also offer new opportunities; for example, smart charging and vehicle-to-grid can be used to support the grid by providing several local and global power- and energy-based services (frequency regulation, peak load sharing, congestion management, reduction of energy curtailment, increase of variable renewable energy source penetration, etc) [142].

Several studies have been performed recently to precisely assess the impact of EV integration into the electrical distribution networks and validate the opportunities offered by EVs. For example, a simulation tool based on a probabilistic three-phase load flow program has been developed and combined with EV usage scenarios simulated by using Monte Carlo techniques [143], by taking into account random variables of EV charging such as the battery state of charge, the starting and stop time of charging, and the EV's location. By using this tool, technical and economic impacts of EV integration on the distribution network were assessed. Studies of the potential opportunities of ancillary services provided by EV were also carried out for low voltage rural and urban networks.

Several strategies for the optimal management and real-time control of EVs in smart cities have been developed (figure 6.39), in order to do the following:

1. Minimize the cost of EV charging (economic criterion)
2. Maximize the use of solar energy to charge EVs (environmental criterion)
3. Reduce the peak load and the power exchange with the main grid (local flexibility)
4. Maximize the contribution of EVs to ancillary services (voltage and frequency control).

6.4.3.5 Major limitations to overcome

Battery deployment is having great success. However, further progress is needed, especially regarding performance, durability, robustness in all climate conditions,

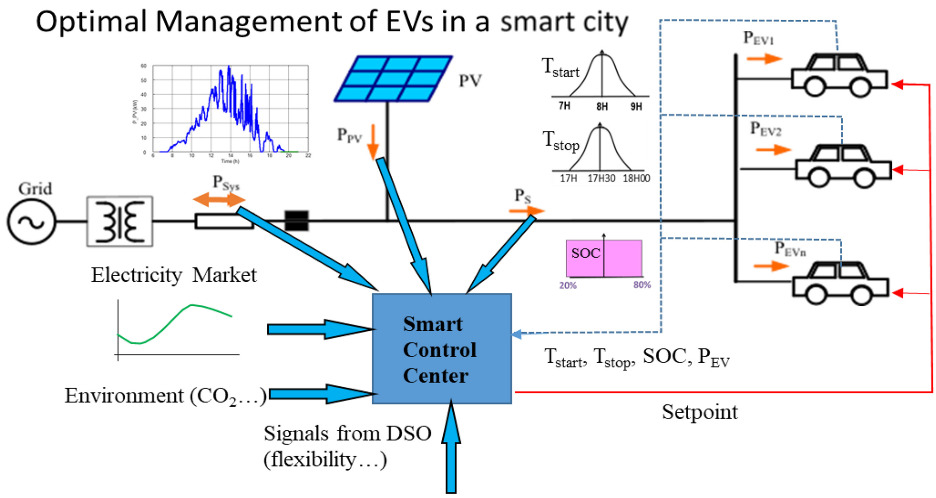


Figure 6.39. Optimal management of EV charging stations in a smart city. DSO: Distribution system operation means a natural or legal person responsible for operating, ensuring the maintenance of, and, if necessary, developing the distribution system in a given area and, where applicable, its interconnections with other systems and for ensuring the long-term ability of the system to meet reasonable demands for the distribution of electricity. SOC: State of charge is the level of charge of an electric battery relative to its capacity.

and safety during operation. In addition, high-performance batteries currently integrate critical materials such as cobalt, which will need to be reduced or substituted in the near future to ensure a large-scale deployment of electrified transport modes. Similarly, hydrogen fuel cells still do not have sufficient lifetime and robustness to be cost competitive, and the amount of critical raw materials such as platinum group metals should be reduced. Hereafter, we describe how physicists can contribute to overcoming those bottlenecks.

6.4.4 Challenges and opportunities for physicists: the role of modeling, simulation, and characterization

Physicists have a crucial role to play in the coming decades to push forward the science and technology of electrochemical power sources such as batteries and hydrogen fuel cells, which are key building blocks for transport electrification and green cities. Improving the performance, durability, and safety of batteries and fuel cells cannot be achieved by relying only on experimental work with conventional trial-and-error methods. It will be necessary to gain a deeper understanding of the complex multiphysics and multiscale phenomena occurring in batteries and fuel cells. This can be achieved by making extensive use of computational simulation in combination with advanced characterization techniques. In the remainder of this section, we describe how physicists can use can use modeling, simulation, and characterization to bring a major contribution to the field of batteries and fuel cells.

In the 250 years that have elapsed since the industrial revolution, scientific discovery has led to a quantitative understanding of matter solidly grounded on the atomistic concept. Ludwig Boltzmann originated a new scientific age by setting the bases of statistical mechanics. At a time when many physicists did not believe in atoms as real entities, Boltzmann's originality was to combine the atomistic hypothesis, with the laws of mechanics that were known at his time plus a new ingredient: statistical reasoning and the math that goes with it. Two of his equations, namely, the definition of entropy and the transport equation, constitute the foundations of equilibrium and nonequilibrium statistical physics. A few decades ago, computers began complementing this approach. Intensive computer modeling and simulation are increasingly able to predict observable macroscopic properties of materials accurately, starting from the atomistic concept and the laws of statistical mechanics.

Concerning batteries or fuel cells, the computational physicist has a lot to contribute to the materials aspects. One way is by finding new materials, screening them for desired properties and for replacing critical materials. This constitutes the goal of high-throughput (HT) approaches, which trade off accuracy for number of compounds and may investigate collections containing more than 1000 compounds [144–149]. Therefore, HT benefits greatly from the aid of machine learning methods like those used in regression analysis and pattern recognition.

However, real materials are more complex than the ideal single crystal. They contain defects and microstructural features that strongly affect their properties, which HT cannot easily handle at present. These investigations of single materials are the realm of multiscale approaches, which encompass the atomic description but extend to length and time scales several orders of magnitude above it. They play a role at least as important as that of HT, allowing us to understand the experiment when none of our expectations is matched by what we measure. Here the physicist starts from the atomic picture but gets to the observable macroscopic properties by means of intermediate equations or models of the intermediate scales. In these investigations too, machine learning methods can be a big help. They help to surmount the computational limitations inherent to *ab initio* techniques, which are often too slow to sample statistically significant numbers of atomic configurations, large system sizes, or long molecular dynamics simulations [150–153].

Both equilibrium and nonequilibrium computational physics are necessary for research into battery and fuel cell materials. Equilibrium modeling includes the prediction of phase diagrams. Since recently, finite temperature phase stability can in many cases, be predicted without adjustable parameters. These are important, for example, in understanding the different phases formed in the cathodes and anode as Li insertion takes place [154]; to estimate vacancy formation probability, which affects whether Li will be able to diffuse in solid battery cathodes; or to evaluate the amount of heat that a thermochemical material can store.

Nonequilibrium materials modeling is also essential to energy storage. On the one hand, there are the linear response conductivities, which are all interrelated via

Onsager's relations [155] to give the transport properties of the material. The electric and ionic conductivities determine how batteries charge and discharge, which is essential for use in small vehicles. In addition, the thermal conductivity is essential to battery safety and plays a fundamental role in the thermal runaway mechanism, which often leads to catastrophic ignition of a battery pack [156]. The dynamic insertion of Li in battery active materials is a complex multiphase process that requires the description of coupled dynamic phase transitions. Their description can benefit from atomic-scale models and simulations to determine both the energies of interactions and the local mobility of Li. In order to be used in larger-scale models, these complex transport mechanisms must be upscaled in the form of simpler and yet relevant laws.

An even more involved aspect of nonequilibrium physics is the prediction of reaction kinetics. Phenomena such as the oxidation of solids contain a lot of complex physics, involving phase transformations, vacancy and interstitial formation, atomic diffusion, and elastic and plastic deformations. Predictive modeling of reaction kinetics thus requires linking the atomic and microstructural descriptions. Not just the macroscopic properties of the material, but also its aging and degradation, depend on the combined physics at these different levels. A promising avenue is to combine *ab initio* and phase field methods. Reaction kinetics is also important for the safety of batteries to understand both the stability of materials and the energy released; the dynamics of this release during their rapid degradation can eventually lead to thermal runaway. An important challenge is the prediction of the chemical composition of the rapid degradation products depending on the composition of the electrodes: active materials, electrolyte, and possible degradation products due to aging.

If the material scale and the atomic scale are of primary importance because, as in any other domain of physics, all the macroscopic properties of interest depend on the atomic scale (local thermodynamic and transport properties), the upper scales are at least as important. Indeed, a battery or a fuel cell is not made of a homogeneous crystal: an electrode is a complex microstructure of heterogeneous multimaterials, a cell is made of a multilayer of electrodes and other components (separator, current collectors, bipolar plates, etc), a module or stack is made of several cells, and so on. Each of these scales presents specific heterogeneities and couplings and represents specific modeling and simulation challenges that must not be underestimated to develop a truly predictive approach to the development of battery and PEMFC technologies. Beyond the development of models at the different intermediate scales of interest, an important challenge is the link between these scales. An important scientific challenge corresponds to the upscaling process: At a given scale, the laws that account for mechanisms known at a lower scale must be determined. The importance of this approach can be illustrated in the durability challenge: To increase the lifetime of batteries and fuel cells requires their management at the system scale such that degradation mechanisms at the microstructure or even atomic scale are not triggered or at least are triggered in only a limited way. Thus, to develop the optimized management strategies, it is necessary to understand and quantify how macroscopic parameters (current, temperature, etc) influence these local phenomena and vice versa.

Because the physicochemical processes at stake are very complex, they cannot be determined *ab initio*. Therefore, they must primarily be determined from experimental observations. Moreover, the models must be fed by quantitative parameters and validated quantitatively. This threefold necessity (observe, characterize, and validate) requires the development of specific experimental characterizations.

Initially, experimental characterizations of fuel cell and battery materials were achieved by electrochemists with the aim to evaluate and improve material performance and ultimately to increase the total system power density and thus the vehicle autonomy. However, the results obtained with pure materials tested separately were often nonrepresentative of their behavior in real systems. Physicists entered the game either implementing high-resolution *ex situ* characterization tools such as transmission electron microscopy [157], x-ray photon spectroscopy (XPS) [158], or time-of-flight secondary ion mass spectroscopy. More recently, they developed *in situ* and *operando* experiments performed on neutron and x-ray sources for studying full systems. Such techniques have led to a better understanding of the transport and degradation mechanisms but require in the case of *operando* observations the development of specific cells and environments adapted to the experimental constraints.

Lithium insertion in cathodes of Li-ion batteries has been studied *operando* by x-ray [159, 160] and neutron diffraction [161] for around 10 years. Recently, using a microfocus beam, it has even been possible to follow the different steps of lithium disinsertion with a micrometer spatial resolution along the thickness of a graphite anode similar to those used in the very large majority of commercial batteries [162]. Specific electrochemical cells are still being developed to study *in situ* and *operando* the lithium insertion mechanisms using different and complementary characterization techniques [163]. XPS could also be used to characterize *operando* the formation of solid electrochemical interphase resulting from the interaction between electrolyte and electrodes [164].

Regarding fuel cells, *operando* characterization was developed, on the one hand, to study electrode reaction kinetics and specifically catalyst efficiency when decreasing platinum loading using EXAFS and, on the other hand, to study cell water management. The water distribution within an operating fuel cell was first characterized globally using neutron imaging [165] and specifically within the electrolyte membranes by small-angle neutron scattering [166]. In both cases, recent significant advances in spatial resolution have permitted researchers to get information on the water content in each electrode as a function of the operating conditions [167].

These results can be advantageously combined with modeling approaches to understand the different mechanisms that are highly coupled and multiscale. The next step will be dedicated to in-depth understanding of the aging mechanisms, which are paramount in evaluating the cost and the sustainability of an energy storage or conversion technology.

6.4.5 Conclusion

In this section we have described how transport electrification in urban areas can contribute to fight climate change and air pollution. We have also presented the role that physicists can take in the coming decades to drive forward the science and technology of batteries and hydrogen fuel cells and hence to make electric vehicles more cost effective, safe, and sustainable.

Although we have focused in this section on urban transport, long-distance transport should not be forgotten. Reaching net-zero greenhouse gas emissions by 2050, as required by the Paris Agreement, means decarbonizing long-distance transport modes such as air transport and shipping. Direct electrification with either batteries or hydrogen fuel cells cannot fully meet the needs of those types of transport modes. Energy carriers with higher volumetric or gravimetric densities are required, which means that low-carbon solutions such as sustainable biofuels (i.e., biofuels made from nonfood biomass) or synthetic fuels (i.e., carbon-based or nitrogen-based fuels, obtained by reacting hydrogen with CO_2 or N_2) [168, 169] should be developed and deployed.

6.5 Environmental safety

Jacob de Boer[1]
[1]Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands

6.5.1 Introduction

When we speak about *environmental safety*, it is good to look first at the definition of *the term environment*. The environment can be defined as 'everything that is around us'. So let's take the place where you live, such as your home. What is around you? Nearby, you may see furniture, carpets, curtains, and electronic equipment, such as computers, mobile phones, and a television. You may see paint and dust, and there is (indoor) air around us. There may be other humans and pets. There may also be unwanted creatures such as insects. And of course, since 2020 we have become painfully aware of the viruses around us. Farther away, there is the outdoor air, and depending on where you live, there may be river or lake water or the sea. There could be meadows and a forest or dunes and a beach. There may also be industry nearby with wastewater and steam and smoke from their stacks. Even farther away are the upper layers of the atmosphere and the oceans. The various environmental compartments also interact, for example, through rain, wind, and temperature that can affect our homes and gardens and ourselves. In other words, the environment is actually relatively complex.

Our safety in our environment is dependent on various factors. There are the aforementioned influences of the environment itself, including biological effects of undesired pests, viruses, flooding, and so on. Even a physical effect such as noise, for example, from a nearby factory or wind turbine, could be felt. There can be also influences of a chemical nature. These can be for example a high nutrient load from wastewater that enters a river, exhaust from industries, contaminants present in our drinking water, house dust, or chemicals that migrate out of mobile phones and computers. In this section, environmental pollution will be discussed with an emphasis on its chemical nature, followed by a discussion of physical, chemical, and political measures to protect ourselves.

Have people ever lived without environmental pollution? If we go back to the Middle Ages or before, there was always personal waste: urine and feces. There were also clothes that people had worn until they were worn off, and remainders of food such as peels. People normally dumped personal waste outside their homes or outside their small villages. Two characteristics of this waste meant that it did not create an environmental problem: The waste was purely natural, and there was a lot of space because areas were never densely populated. When cities grew, environmental problems also grew. Cities began to smell due to putrefaction. This caused illnesses because bacteria started to grow and spread around. In Amsterdam, the drinking water, which initially came straight from the canals and later from the nearby river Vecht, was so bad that people decided to drink only beer (with 1%–2% alcohol) to survive. For the beer, water was used from the nearby lake Haarlemmermeer. In 1845, Amsterdam decided to install an entire new drinking

Figure 6.40. The Dutch dunes provide an excellent natural filter to produce clean drinking water.

water system with natural filtration in the Dutch dunes; it opened on 12 December 1853. This system now produces 70 000 000 m^3 of pure drinking water per year [170] (figure 6.40).

The example of the Amsterdam's drinking water shows that the initial environmental problems were of a biological nature. The industrial revolution that took place in the 1800s profoundly changed this situation. Suddenly, it was possible to synthesize chemicals for all sort of purposes. Those chemicals were very useful, such as medicines, dyes, fertilizers, and eventually plastics. Unfortunately, waste streams were created as well. These Included wastewater containing intermediates from production, polluted vapours that were released into the air, and the products themselves which did not degrade after use. Cars were produced, which produced exhaust, and trains and airplanes with comparable or more exhaust. Ships using bunker oil started to pollute harbors and ocean air. The Haber–Bosch process was invented in 1909 and created an almost endless source of fertilizer (and explosives) [171]. Due to the use of these artificial fertilizers—more than 100 000 000 metric tons annually—more food could be produced, and it was possible to feed many more people, which stimulated the growth of the world population. However, pumping massive amounts of synthetic nitrogen products into the food chain also created massive algal blooms in waterways and the proliferation of ocean dead zones. The concerns about synthetic nitrogenous fertilizer now have a greater immediacy than ever [171]. Meanwhile, the combination of a world population of more than seven billion people and a continuously growing production of chemicals have created environmental pollution at a global scale.

Unfortunately, problems with environmental safety are not the only ones caused by the strong population growth. There are serious problems with food and water scarcity. Energy production may be one of the greatest threats at the moment. If we are not able to make much more use of alternative energy resources such as solar and wind energy or nuclear energy, global warming will continue to a level that causes serious effects on many people, through flooding, droughts, hurricanes, or the temperature itself. Clearly, the problems with the COVID-19 virus (and other viruses such as other SARS viruses and MERS viruses) are also related to the dense world population. These problems may be more threatening for humankind than chemical pollution. However, the combination of the exponential growth of the world population and the industrial revolution has resulted in unprecedented levels of chemical pollution. This section will focus in particular on this aspect.

6.5.2 Current developments in production and restriction of chemicals

Between 1965 and 2006, global production of human-made chemicals increased exponentially. In 1965, 0.2 million substances were registered by Chemical Abstracts Service, and 88 million were registered in 2006 [172]. This exponential increase continues every day, with a staggering 10 million new compounds being produced each year. This is more than 1000 chemicals every hour, or 17 per minute [173]. Certainly, not all of the new chemicals find their way to commercial production and use. However, the growth in the chemical production is exponential. The United Nations (UN) Global Chemicals Outlook II predicts a doubling of the total global volume of chemicals used in the period 2020–30 [174]. There are currently no predictions for the period after 2030. According to US Environmental Protection Agency (EPA), there are approximately 85 000 chemicals in commerce in the US only. This number is, however, debated. Because of many duplications in that list, the real number would be around 35 000. Given that the number of evaluations taking place by the US EPA [175] and by the EU REACH program (Registration, Evaluation, Authorisation and Administration of Chemicals) are no more than a few hundred, it is clear that control is falling behind. To find our way in this chemical labyrinth, the most relevant categories of environmental pollutants will be discussed. This selection of chemicals is made based on their harmful properties in combination with their production volumes. The latter is important because, according to the old but still very relevant adage of Paracelsus, the dose determines the effect of a certain compound. Therefore, chemicals produced in low volumes will not be very important as regards environmental risk, except in local situations.

6.5.2.1 Persistent organic pollutants

To determine whether properties are harmful, the PBT concept is one of the most used tools. PBT stands for persistent, bioaccumulating, and toxic. A compound with this combination of properties would be one of the first to be categorized as undesirable. The PBT concept also underlines the definition of POPs: persistent organic pollutants. This definition was used by the UN in the Stockholm

Table 6.2. Current decision criteria for persistence, bioaccumulation, and toxicity of chemicals [176].

Compartment	Persistence half-life (d)	Bioaccumulation factor	Environmental toxicity, NOEC (mg L^{-1})	Human toxicology
In air				
In water	>40[a], >60[b], >60[c]			
In soil	>120, 180[c]			
In sediment	>120[a], >180[b], >180[c]			
Biota/water		>2000, >5000[d]	< 0.01[a], <0.01[b]	
Humans				CRM or chronic toxicity

Notes:[a] Freshwater.
[b] Marine water.
[c] Very persistent.
[d] Very bioaccumulative.
CRM: carcinogenic (category 1 or 2), mutagenic (category 1 or 2), or toxic for reproduction (category 1, 2, or 3).
NOEC: no-observed-effect concentration.

Convention, which was set up to ban all POPs worldwide. The criteria for persistence, bioaccumulation, and toxicity have developed through the years. Especially for toxicity, criteria may change if new study results become available. In general, toxicity criteria will therefore become more strict. Table 6.2 shows the current decision criteria [176].

Given the very wide acceptance of the UN Stockholm Convention, it is fair to say that there is a global agreement on banning POPs. The initial list of POPs contained 12 substances. Unfortunately, the list had grown since the Convention entered into force in 2004, instead of shrinking. New discoveries and better toxicological insights have led to adding more substances to this Convention. At the moment there are 35 substances listed as POPs. Although it is 17 years since the Stockholm Convention came into force, no substances have been cleared. There are still many illegal dump sites with DDT or other pesticides, many old transformers containing polychlorinated biphenyls (PCBs), and so on. On the positive side, production of almost all POPs has stopped, although, in particular on the more recently added POPs, several countries have asked for and received exemptions, often for economic reasons. Because so many countries are involved, the Stockholm Convention is a rather bureaucratic organization in which progress is slow. The main problem is countries' capacity for the destruction of chemicals. The initial POP list contained PCBs, dioxins, and a number of organochlorine pesticides such as DDT, dieldrin, heptachlor, mirex, and toxaphene. The list was extended with endosulfan, several brominated flame retardants, perfluorinated alkyl substances (PFAS), and a few others. All POPs are halogenated substances. Whereas initially chlorinated compounds were considered, it appeared later that brominated and fluorinated compounds also behave as POPs. Especially, the carbon-fluorine bond is extremely strong. The persistence of the PFAS is therefore very high. An article in the *Washington Post* coined the term ‘forever chemicals’ because once produced, they will be present in nature for very long times, leaving a problematic heritage for future generations [177].

Positive aspects of the Stockholm Convention include creation of awareness and production of environmental data and of course the termination of production across the entire world. Global collaboration is essential to reduce the presence of these substances. Once they are released in the environment, they are distributed worldwide, mainly due to the so-called grasshopper effect [178]. They evaporate, precipitate again with rain or snow, evaporate again, precipitate, and so on. The direction is away from the equator towards the Arctic and Antarctica. Because production was mainly in the northern hemisphere, the Arctic is a very vulnerable area for POPs. This long-range transport is the fourth pillar of the Convention. Potential for long range transport is considered in the following circumstances:

1. When a chemical can be found in concerning concentrations at locations distant from the sources of its release.
2. When monitoring data show that long-range environmental transport of a chemical may have occurred via air, water, or migratory species.
3. When environmental fate properties and/or model results demonstrate that the chemical has a potential for long-range environmental transport through air, water, or migratory species. For a chemical that migrates significantly through the air, its half-life in air should be greater than two days [179].

The POPs in the Stockholm Convention are listed in three annexes, with different requirements [179]. For the initial 12 POPs (in annex A), all production and new uses should have ended in 2009, but exemptions were made for POPs that were added later to the list, such as hexabromocyclododecane, pentachlorophenol, chlorinated paraffins, lindane, endosulfan, and polybrominated diphenyl ethers. For the two POPs listed in annex B, DDT and perfluorooctane sulfonate (PFOS), no safe and affordable alternatives are yet available; therefore, they can still be produced or used for certain purposes. Notification to the Secretariat of production or use of chemicals under annex B is required. For DDT, the only acceptable purpose is disease vector control in some specific countries. For PFOS, research on alternatives is ongoing, but at this time, the production of Teflon cannot be achieved without fluorinated intermediates (figure 6.41). For use in firefighting foam and for coatings on outdoor wear, alternatives have been proposed [180, 181]. PFAS is a large group of chemicals, consisting of more than 5000 varieties with new ones being regularly added. This often leads to so-called regrettable substitution, which means that the alternative is not much better or even worse than the chemical that was replaced. One example is the substitution of perfluorinated octanoic acid (PFOA or C8) by Gen X (figure 6.42). In Gen X an oxygen atom is present which makes the compounds much less bioaccumulative. The half-life in the human body is

Figure 6.41. PFAS, often present in pizza boxes and other fast-food packaging.

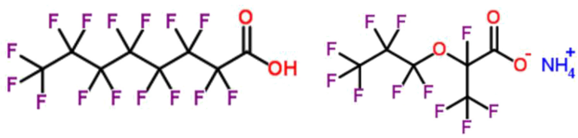


Figure 6.42. PFOA (left), used as intermediary in the production of Teflon, was replaced by Gen X (right), an example of regrettable substitution.

around 24 h instead of approximately 4–6 years for PFOA. However, Gen X dissolves much better in water, making it a much more mobile chemical. When preparing drinking water from surface water contaminated with PFOA, drinking water companies have great difficulties removing Gen X, so it arrives in drinking water, causing a daily uptake by consumers [182].

The lesson from this example is that a ban makes sense only when it includes the entire group of PFAS. Initiatives to realize this are currently being taken in Europe. Such a ban may include exemptions for essential use [183]. This concept addresses the use of chemicals for which no alternatives are yet available but which offer essential benefits, for example in medical applications. In the same way, within the group of halogenated flame retardants, one flame retardant is being replaced by another without making progress from an environmental point of view. The only solution is a ban on an entire group of compounds (figure 6.43) [184].

Chemicals in annex C are formed and released unintentionally during the production of other compounds or in thermal processes, such as during the incineration of waste. In order to prevent formation of these POPs, countries need

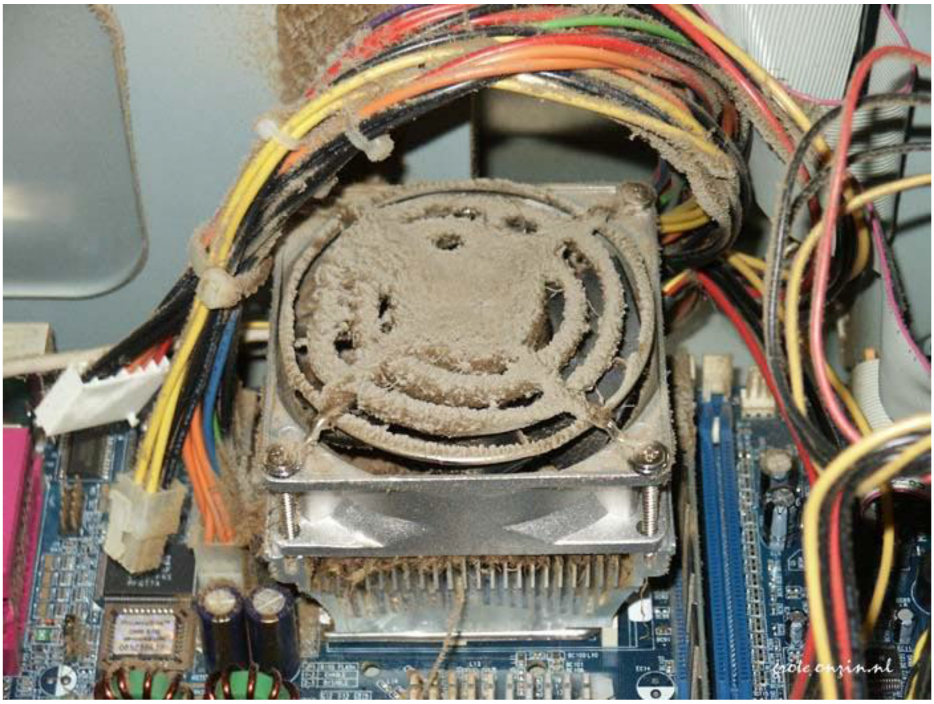


Figure 6.43. House dust contains high amounts of toxic flame retardants.

to develop release inventories, apply the best available techniques and promote the best environmental practices, and report progress every 5 years. Compounds in annex C comprise polychlorinated dibenzo-*p*-dioxins, polychlorinated dibenzofurans, and pentachlorobenzene. Some chemicals, such as polychlorinated naphthalenes, hexachlorobenzene, and PCBs are in both annex A and annex C [179].

6.5.2.2 Endocrine disruptors

Endocrine-active compounds or endocrine disruptors (EDCs) are chemicals that disturb the hormonal process in animals and humans. EDCs are distinguished in synthetic (S-) EDCs and natural (N-) EDCs [185]. The group has a high diversity and runs across other categories of chemicals, such as the aforementioned POPs. Natural hormones such as estradiol and progesterone are also included. The function of the endocrine system is strictly regulated, involving the hypothalamic–pituitary–gonad axis. The endocrine system can be modulated in two basic ways. The first is by agonists or antagonists of the respective estrogen and androgen receptors. Compounds that mimic, to some extent, the behavior of hormones can disturb the hormonal balance. A debate is going on about which compounds should be considered as EDC as which should not and also whether the N-EDCs may be having a much stronger effect than the S-EDCs [186]. Isoflavones from soy-based food and coumestans in certain vegetables are examples of N-EDCs that contribute substantially to daily human consumption. On the other hand, reports suggest an association between S-EDCs and obesity, for example, for parabens, phthalates, and bis-phenol-A. Another diverse group of bioactive chemicals receiving comparatively little attention as potential environmental pollutants includes pharmaceuticals and personal care products, both human and veterinary, including prescription drugs, diagnostic agents, sunscreen agents, fragrances, and many others. These compounds and their bioactive metabolites can be continually introduced to the aquatic environment as complex mixtures via a number of routes but primarily by both untreated and treated sewage. Many also show EDC activity [187].

6.5.2.3 Pesticides

Some chlorinated pesticides, such as DDT, dieldrin, toxaphene, chlordanes, and lindane, and their metabolites have been categorized under Stockholm Convention as POPs that should be phased out [179]. There are, however, numerous categories of pesticides—euphemistically called plant protection products—that have been developed to replace the chlorinated pesticides or for specific purposes. The enormous growth of the world population and the associated demand for food has greatly stimulated the use of pesticides. Without the use of pesticides, it is impossible to maintain even the current high food production for the entire world. That does not take away from the fact that many pesticides are toxic for the environment and humans. There are painful stories of casualties in Africa due to improper use of pesticides. This is often related to the lack of knowledge about the application of pesticides, personal protection, and proper storage.

Apart from direct intoxication, many pesticides have chronic effects. Debates are taking places on the admission of certain frequently used pesticides such as glyphosphate. In 2017, the European Commission reapproved the use of glyphosphate, the world's most widely used active ingredient in herbicides, in spite of a heavy debate [188]. These authors called for a new system of categorization of pesticides in which regulatory decision making takes into account not only the technical evidence on safety but also the societal context. Another group of heavily debated pesticides are the neonicotinoides, which are suspected to cause detrimental effects on pollinators such as honey bees. Thirty-five percent of the world's food crop production depends on pollinators [189], so although pesticides are essential for the world food production, this type of pesticide may eventually be counterproductive. In addition to the example mentioned, there is a very large collection of pesticides that are used worldwide, from easily degradable to persistent, and from relatively nontoxic to toxic. New pesticides are regularly introduced to the market, and food control agencies have a hard time continuously checking for residues.

6.5.2.4 Nutrients and fertilizers

Nutrients are essential for growth of plants, animals, and humans. However, nutrients are also excreted by animals and humans, which, when population density is high, can create environmental problems. On the other hand, to meet the global demand for food, natural nutrients fall short, so fertilizers are being added to agricultural soil to enhance food production. The enormous pressure created by the current large world population and the heavy use of fertilizers seem to have pushed the nutrient system out of balance [171]. Too large an amount of nitrates and ammonium on land may cause damage to forests. Too large an amount of phosphates and nitrates in water causes eutrophication and possibly undesired plankton blooms in our seas. In Western countries, sewage treatment systems have been built that remove most nitrates from wastewater. Experimental systems may even remove phosphates. However, in most developing countries, the situation is completely different. Sanitation is sometimes completely unavailable. Nitrous and sulfuric vapours from industries can cause acid rain and air pollution with serious risks for human health. The World Health Organization estimates that air pollution causes the annual premature death of two million people worldwide [190]. To put this into perspective, this is twofold higher than the current annual number of COVID-19 victims.

6.5.2.5 Polycyclic aromatic hydrocarbons

Polycyclic aromatic hydrocarbons (PAHs) are carcinogenic compounds that occur naturally. They may also be produced unintentionally during combustion of fossil fuels. In spite of their natural character, they can cause serious environmental pollution. Traffic, especially cars and airplanes, and heavy industry are the main producers of PAHs that are being formed during combustion processes in engines. PAHs are semipersistent, having the ability to stay for a very long time in environmental compartments such as soil and sediments. In organisms, PAHs are subject to metabolism processes, which does not mean that they are less harmful. Benz(a)pyrene in particular is highly carcinogenic. PAHs are also easily generated in open fireplaces and during barbecues, making houses and gardens sometimes a critical environment. PAHs occur in rubber, and through abrasion from car tires, they reach the environment near roads. The offshore industry and steel factories are other sources of PAHs. Also, they occur in crumb rubber, which is still used on football pitches. In addition to PAHs, methylated, oxygenated and hydroxyl PAHs also occur [191].

6.5.2.6 Synthetic dyes

The accelerating textile industry requires high volumes of dyes. The dye manufacturing industry represents a relatively small part of the overall chemical industries. The worldwide production of dyes is nearly 800 000 tons per year [192]. About 10%–15% of synthetic dyes are lost during different processes of textile industry. Textile factories, such as for example in India, struggle to get rid of the dye waste from their factories in an environment-friendly way. There are more than 10 000 dyes used in textile manufacturing alone, nearly 70% being azo dyes which are complex in structure and synthetic in nature. Textile industries produce large amounts of liquid waste. These textile effluents contain organic and inorganic compounds [193]. During the dyeing processes, not all dyes that are applied to the fabrics are fixed on them and there is always a portion of these dyes that remains unfixed to the fabrics and gets washed out. These unfixed dyes are found in high concentrations in textile effluents. Effluents are rich in dyes and chemicals, some of which are nonbiodegradable and carcinogenic and pose a major threat to health and the environment. Several primary, secondary and tertiary treatment processes like flocculation, trickling filters and electrodialysis have been used to treat these effluents. However, these treatments are not found effective for the removal of all dyes and chemicals used. The usage of cotton has been increasing constantly throughout the past century. Cotton fibers are mainly dyed using azo dyes, which are one of the largest groups of synthetic colorants used in the industry [194]. Azo dyes are difficult to degrade by the current conventional treatment processes. Advanced oxidation processes (AOPs) can be used to treat dyes. The main advantage of AOPs over the other treatment processes is its pronounced destructive nature, which results in the mineralization of organic contaminants present in wastewater [193].

6.5.2.7 Greenhouse gases and CFCs

At the moment, global warming, caused by greenhouse gases mainly, is obviously, one of the greatest threats for humans. Although three gases are always mentioned as the most important greenhouse gases, carbon dioxide, methane and nitrous oxide, water vapor is the most important actor in causing greenhouse effects, twice as powerful as carbon dioxide [195]. In addition, there are a few other gases causing global warming to a minor extent, such as volatile halogenated compounds (see below). Global warming is a rather complex process, not only related to industrial activity. Carbon dioxide, methane, nitrous oxide and water vapor are being produced by both natural and anthropogenic sources. Carbon dioxide emissions are for example large on a human scale but small compared to natural fluxes due to photosynthesis, and uptake and release into and from oceans. Methane is more related to anthropogenic sources such as mining, petroleum industry, coal combustion, biomass burning and domestic sewage treatment, than to natural sources. Nitrous oxide is mainly coming from natural sources but use of fertilizers, biomass burning and fossil fuel combustion also contribute.

Chlorofluorocarbons (CFCs) are stable, synthetic, halogenated alkanes, developed in the early 1930s as safe alternatives to ammonia and sulphur dioxide in refrigeration. Production of CFC-12 (dichlorodifluoromethane, CF2CIz) began in 1931 followed by CFC-11 (trichlorofluoromethane, CFCI3) in 1936 [196]. Many other CFC compounds have since been produced, most notably CFC-113 (trichlorotrifluoroethane, C2F3CI3). CFCs are nonflammable, noncorrosive, nonexplosive, very low in toxicity, and have physical properties conducive to a wide range of industrial and refrigerant applications. Many of these CFCs also have a, relatively small, effect on global warming. They are, however, more known from causing damage to the ozone layer, allowing harmful solar UV radiation reach the Earth's surface. The Montreal Protocol on Substances that Deplete the Ozone Layer was adopted in 1987. This protocol is often described as a unique example of a global agreement that successfully averted an environmental crisis [197]. It is still the only UN treaty ratified by all 197 member states. The first paper on ozone layer depletion by Molina and Rowland [198] was published in Nature in 1974. In 1995, Molina and Rowland, together with Crutzen, who focused on the effect of nitrous oxides on the ozone layer [199], received the Nobel prize for their discovery of these effect on the ozone layer and their efforts leading to the Montreal protocol. An overview of how basic research led to a worldwide agreed protocol for control of the harmful CFC's was given in the Nobel lecture by Rowland [200]. Scientists report the level of UV radiation could have quadrupled since the 1980s, whereas now UV radiation has stayed constant during the last 20 years [201]. Nevertheless, attention remains needed. Currently, in the atmosphere volumes of hydrofluorocarbons (HFCs), which have limited effect on the ozone layer, but are potent greenhouse gases, are growing rapidly, with an average rate of 1.6 ppt (parts per trillion) per year between 2012 and 2016 [202]. In 2016 the phase-down of production and consumption of HFCs was added to the Montreal protocol in the Kigali Amendment [203].

6.5.2.8 Trace metals

Although of natural origin, trace metals were one of the first compound classes to be considered as environmental pollutants. In the Minamata incident, waste from a factory producing mercury sulfate as catalyser was released into Minamata Bay (Japan), where methylmercury accumulated in fish. Fish-eating cats were the first victims, followed by birds falling from the sky and dying dogs. Fishermen caught fish with two or three heads. Finally, more than 1700 people died, including children [204]. Mercury binds to proteins that support the nervous system. Mining and steel industries initially caused local and later regional and global elevated levels of mercury in particular, but also of other trace metals such as copper, arsenic, cadmium, and lead. Nowadays, trace metals are still threatening the environment, but globally there is a lot more attention for this type of pollution. The Minamata Convention of the United Nations monitors mercury flows worldwide. Many countries have monitoring programs in which trace metals are being analyzed in water, sediments, and fish. One of the bigger remaining problems is the high level of arsenic, often of natural origin, that pollutes drinking water in many areas in India and Pakistan and some other developing countries. A shift in the use of metals is seen towards use of rare earth elements (REE), which are used in magnets and military systems. Recycling of REE would be a possible solution, but so far, no reliable and sustainable techniques are available [205]. These REE are often mined from areas that are highly vulnerable. Cobalt is another metal that indirectly threatens the environment. Cobalt is used in batteries of mobile phones, and stocks of cobalt are available in areas in Congo where the highly threatened mountain gorillas have their habitat [206]. DR Congo is a the top producer of cobalt with 40% of world production and one-third of world reserves.

6.5.2.9 Radionuclides

In the human perception the risk of radionuclides is always considered very high. This may be related to the two atomic bombs that were dropped on Hiroshima and Nagasaki in Japan and the nuclear tests during the Cold War. Also, serious accidents such as those at the nuclear power plants at Chernobyl and Fukushima, which have contaminated parts of the world, have contributed to the general fear. Indeed radionuclides can be health threatening. Some of them have very long half-lives, which can cause a risk for many generations. On the other hand, aside from these accidents, the level of pollution from nuclear reactors is extremely low. It does not threaten environmental or human lives and is often far below natural background radiation as experienced in airplanes or high in the mountains. Benefits of radionuclides and radiation should not be forgotten. Many radionuclides are essential in the treatment of cancer. X-rays are pivotal in detecting fractures and many other health-threatening impairments. Medical radiation exposures are, therefore, often life-saving, although at the same time they constitute the majority of exposures in terms of number of persons exposed and higher dose rates. In the environment, a somewhat higher risk may be found in the production of phosphate fertilizers, which often leads to pollution with the alpha-radioactive lead-210 and polonium-210 [207, 208]. These compounds contribute to the occurrence of cancer due to smoking, because tobacco plants accumulate these compounds from the fertilizers.

6.5.2.10 Plastics

Although pollution of beaches has been well known since the 1960s, scientists have ignored this topic for many years. Possibly it was seen as a problem that was so easy to avoid through human behavior that no one bothered to approach it in a scientific way. It was not until 1980 that reports started to appear in the literature on the risks of macroplastics and microplastics [209]. Large plastics field floating around in gyros in the Pacific Ocean were reported [210]. Beach monitoring of litter was started, and several initiatives involving plastic monitoring and suggestions for cleaning were observed. It appeared that microplastics with a size range of 0.1–5 mm were a large part of the problem. Microplastics can be formed, though slowly, from weathering of larger pieces of macroplastic. Spills in harbor of commodities for plastic production is another source. Microplastics are also present in many products such as toothpaste and cosmetics, and reach the rivers and sea because they pass through sewage treatment systems.

Macroplastics and microplastics are of serious concern for the environment [211] (figure 6.48). Macroplastics cause risks for turtles, which can be strangled, and for birds that consume plastics and stop feeling hungry, which eventually causes starvation because they do not take in their normal diet anymore. Microplastics can shade light and, therefore, affect the growth of algae and benthic organisms. They enter the gastrointestinal tracts of mussels, other shellfish species, and of fish. Some scientists claim that microplastics are also harmful for humans, causing inflammations. More research is needed on this subject to provide evidence. The analysis of microplastics is difficult because there is not only the chemical component, that is, the various types of the plastics, but also a physical component that can cause damage to organisms. Microplastics have many different shapes, and fibers, for example, from clothes, are included in this group. In addition, chemical contaminants in water can adsorb to the plastics and cause an additional risk for organisms. Nanoplastics are also being studied, and some scientists are worried about harmful effects because, due to their smaller size, they could more easily enter cells or organisms. This research is, however, still in its infancy.

6.5.2.11 Fine particles, NO_x and SO_x

Although estimates vary, there is no doubt about the high mortality caused by air pollution. According to Landrigan [212], polluted air was responsible in 2015 for 6.4 million deaths worldwide: 2.8 million from household air pollution and 4.2 million from ambient air pollution. Lelieveld *et al* [213] estimated the number of death related to air pollution at 3.3 million per year, mainly caused by exposure to fine particles, smaller than 2.5 mm (PM2.5). Lelieveld *et al* [213] mention that agricultural sources are the second-largest contributor to global mortality because releases of fine particles and ammonia from livestock and fertilizers lead to atmospheric formation of ammonium nitrate and sulfate particles. Agricultural sources are leading sources of mortality in the eastern US, Russia, Turkey, Korea, Japan, and Europe, contributing to more than 40% of deaths in many European countries. Coal-fired power plants and steel industries are other dominant sources of air pollution. Apart from mercury (see section 6.5.2.8), they produce high levels of fine particles and significant volumes of NO_x and SO_x vapours. Fine particles from the steel industry pose a double risk, as they have absorbed many different metals (figure 6.44). Traffic is especially causing high levels of fine particles and NO_x exhaust. Some densely populated cities in countries such as China and India have permanently too high PM2.5 levels (figure 6.45). Bronchitis, asthma, strokes, cardiovascular diseases. and lung cancer are examples of diseases caused by air pollution, but it has also been suggested that air pollution is an important although not yet quantified risk factor for neurodevelopmental disorders in children and neurodegenerative diseases in adults [212]. Lelieveld *et al* [213] project a doubling of mortality from air pollution by 2050 on the basis of projected rates of increase in pollution and population levels. This projection should obviously sound alarm bells for public health agencies around the world.

Figure 6.44. Dumpsite in nature, Kerala, India, including a lot of plastic waste.

Figure 6.45. Steel production releases lots of toxic fumes, fine particles and CO_2.

6.5.3 Challenges and opportunities towards 2050

6.5.3.1 Challenges: How to cope with such a load of chemicals?

As mentioned earlier in this section, the UN Global Chemicals Outlook II expects a 6.6 trillion Euro/year market for chemicals in 2030, double the amount in 2020. That is obviously not the end of the growth. Will our world be able to cope with so many chemicals? There is no doubt about the benefits of chemicals for humankind. However, the same report predicts that hazardous chemicals will continue to be released in large quantities. International treaties and voluntary instruments may reduce the risks of some chemicals and wastes, but progress has been uneven, and implementation gaps will remain [174]. It is, therefore, obvious that more is needed than just hoping for the better. Small improvements in the current situation will not help. Similar to the increased consciousness about global warming, there is a need for a world plan for chemical pollution. Today most lactating women in the world still have DDT in their breast milk. Most people in the northern hemisphere have PFAS levels in their blood close to or over the safe levels advised by EFSA. Each year 1.6 million individuals are affected by air pollution. If we do not want to spoil this world for future generations, our ways to deal with chemicals need some fundamental changes. There are solutions that involve using better techniques, investing in innovation, and changing our policies.

6.5.3.2 Better products

To produce high-quality, useful products, many chemical industries produce substantial amounts of waste. Waste often leads to local, regional, or even global pollution. However, some industries make chemical compounds, which are chemical waste themselves as soon as they are produced. Getting rid of those substances often requires a very special type of treatment, such as burning at very high temperatures. A lot would therefore be gained if we could refrain from producing persistent, bioaccumulating, and toxic compounds. Some frameworks have been erected to realize this goal. In Europe the REACH program should ensure that such dangerous compounds are not made any longer. In the US, the EPA is trying to reach the same goal with the Toxic Substance Control Act. Unfortunately, these programs cannot keep pace with the introduction of new chemicals [174]. The only solution is to test chemicals *before* they are produced and used. The pharmaceuticals industry needs to prove the safety of drugs before they can be produced for the market. It has been a big mistake not to test bulk chemicals before production can start. That has caused the aforementioned situations with DDT in human milk and PFAS in human blood, which are basically experiments on a global scale. The chemical industry will always protest and claim that costs will be much too high. A level playing field is therefore needed globally to realize this. Until a proper system of testing has been installed, the precautionary principle needs to be applied [214]. This system prescribes that if a certain chemical substance, based on its structure and on modeling outcomes, may likely cause damage to the environment and/or to human health, it should not be produced and used. Industries may claim that alternatives are not available. However, we heard that argument many times before, such as when PCBs had to be phased out. Alternatives were found. The same argument was used when some brominated flame retardants were phased out. Again there were alternatives. The European research project ENFIRO showed that within 3 years a series of environmentally friendly flame retardants could be selected and produced for major applications [215]. Some industries do produce environmentally friendly and environmentally unfriendly compounds at the same time, showing their good will. Shareholders then press for keep selling the unfriendly alternatives, as profit is often better for those substances. Which chemical industry will stand up and decide to produce only safe chemicals?

Apart from safe alternative chemicals, there is also mileage in thinking about the purpose of using specific chemicals. Flame retardants may be necessary to delay ignition of plastics that are heated, such as in computers and television sets. But do they need to be used in paint on the walls of houses or in flags? Are we not able to find alternatives such as glass or wood to the many plastic furniture items we have in our homes? There are many ways to reduce the use of flame retardants without elevating the risk of fire. The concept of essential use [183] may help us also in the world of flame retardants to decide which ones are really necessary and which ones we can do without.

6.5.3.3 Better processes

The environment has a self-cleaning capacity. The system is able to transform and metabolize chemical compounds in a rather effective way. UV light, bacteria, temperature, and so on will cause chemicals to be converted into water and carbon dioxide. This wonderful process works fine as long as the system is not overwhelmed and as long as no chemicals that are insensitive to UV light, bacterial degradation, and so on are being discharged. It is amazing to see that during the last half century we have entirely ignored the self-cleaning capacity of the environmental system. Certainly locally and also regionally but even globally we have released so many persistent and toxic compounds into the environment that some of them will remain there for generations to come. Some pharmaceuticals in fish farming, for example, have been released in such high quantities that a phenomenon called antibiotic resistance is taking place, that is, the entire environment has been made sterile. Due to their extreme persistence the PFAS have been called *forever chemicals*, because we do not know when they will ever disappear from our systems. Microplastics and macroplastics will float around (and sink to the sea floor) for ages to come, simply because the self-cleaning capacity of the environment is close to zero for the large polymeric molecules.

The solution is actually rather simple: Go back to respecting the self-cleaning capacity of our world. That means basically, taking care of our waste, as we should do already as citizens. Where citizens are supposed to collect their waste and offer it to the garbage collection service, industry should act in a similar way. Just don't throw the waste into the environment (figure 6.46), either into the air or into the water. Factories need to destroy their waste on their premises or send it to a service that can handle it properly. A big Teflon plant in the Netherlands has discharged tens of thousands of kilograms of PFAS into the rivers of the Netherlands. This was not illegal; the authorities had provided the plant operator with permits to do so. When we discovered

Figure 6.46. The extremely densely populated city of Jaipur (India).

that PFAS could be found in fruit in gardens around the plant [182], the owners claimed that production would be impossible without such discharge. However, after more pressure, when building companies discovered they could not build anymore because most locations exceed the maximum tolerance levels for PFAS, suddenly there appeared to be options, such as by installing carbon filters and replacing them with clean ones from time to time and sending the polluted filters to be burned at high temperature. Better filter installations, sewage treatment systems, absorption materials, and so on will help to solve these issues. This is not a matter for industry alone. Even in 2021, in the Netherlands a permit was given to a company to discharge 12 000 kg yr^{-1} of microplastics into the river Meuse, even though everyone is aware of the risks of microplastics and efforts are being made to clean microplastics from rivers and seas. It looks as if public services are mutually disconnected. When Gen X was introduced as an alternative to PFOA for producing Teflon and a permit for discharge was provided to the industry, the drinking water service situated near and downstream from the Teflon factory was not informed about this new compound. Consequently, for years they tried hard to identify the unknown peaks they observed with their instruments. One simple phone call would have been enough to avoid all these efforts. Apart from that, Gen X appeared to be a typical regrettable substitute (figure 6.42).

It is unrealistic to strive for a world without plastic or steel. Even pesticides are needed to ensure sufficient food production for the large global population. However, we need technical innovation, including better filters, clever adsorption procedures, use of safer and cleaner materials, less polluting energy systems such as using hydrogen as fuel for cars and airplanes, and solar energy to reduce air pollution. Physical and chemical designing should be much more focused on environmental and human safety, rather than on making profit.

6.5.3.4 International developments

In 2016, the UN put forth sustainable development goals (SDGs), comprising 17 goals and 169 targets [216]. Less known is that these SDGs also contain 42 targets that focus on means of implementation. As regards technology, there is a focus on transferring technologies from developed to less developed countries. This would certainly assist in developing environmentally friendly production systems and increase the safety of farmers. Although the SDGs have been criticized, for instance, for a lack of interlinkages and interdependencies among goals, the SDGs do give strong guidelines on a global scale [217]. It will be up to authorities and governments to translate these SDGs into laws and international agreements.

The EU has announced substantial action with its Green Deal on a climate-neutral continent. In the wake of the Green Deal, other plans have been introduced, such as the EU Chemicals Strategy for Sustainability, presented on14 October 2020; the Zero Pollution Ambition; the European Industry Strategy; and the EU Hydrogen Strategy. These ambitious plans strive for a strong reduction of pollution and the phase-out of persistent chemical groups such as PFAS and emphasize the need for technical innovation.

In the US, the CSS Strategic Research Action Plan (StRAP) outlines the EPA's research on the development of innovative science to support safe and sustainable selection, design, and use of chemicals and materials required to promote human and environmental health and sustainability [218]. The StRAP will, among other things, provide a chemical safety informatics infrastructure to support decision makers, introduce high-throughput hazard and exposure approaches to fulfill data needs, and consider sensitive populations and life stages in chemical safety evaluations. StRAP will build a broader understanding of biology, chemical toxicity, and exposure while providing more rapid, cost-effective approaches that protect human health and valued ecological resources and services.

The most influential country as regards environmental chemical safety is China. According to the UN Global Chemicals Outlook II, in 2030 China will have a 49% share of the chemicals market [174]. Few global actions on environmental safety will therefore succeed without the participation of China. The EU and China have developed the EU–China 2020 Strategic Agenda for Cooperation [92]. This document contains one paragraph (V.9) on the environment, which involves cooperating on tackling air, water, and soil pollution and chemical pollution; sustainable waste management and resource efficiency within consumption and production; and environmental pollution emergency action. At this stage, it is somewhat unclear how powerful this document will be. Apart from common concern about environmental issues there are obvious competitive issues, such as shortages of rare earth metals and other chemicals which can be found in one region and not in the other. On 8 January 2019, China published the draft of its new Regulation on the Environmental Risk Assessment and Control of Chemical Substances (Consultative Draft) [219]. The new regulation intends to govern the environmental risk assessment and control of any chemical substance. It is the first chemical control legislation proposed by China which covers both existing and new chemical substances. Given also the Chinese development activities in other countries, particularly in Africa, the Chinese influence on global environmental safety in the coming decades should never be underestimated.

6.5.4 Conclusions and outlook

Within the present decade our densely populated world will see a vast expansion of chemicals that will serve many individuals but at the same threaten our environmental safety and our own health. Given the very short notice of this development, immediate action is required. The EU has set an ambitious example with a set of plans including the Chemical Strategy for Sustainability. Comparable action plans in the US and China and collaboration to meet the UN SDGs will be essential. Environmental safety is very much related to the use and policies of dealing with chemicals. However, technical innovation is required to handle these chemicals in a safe way, produce new and safer alternatives, safely collect and handle waste, and produce energy without negative effects on the environment. This is a huge challenge. These changes will not be easy. They will need to overcome strong opposition by powerful vested interests [212]. Much will depend on success in achieving a circular economy. This will be needed not only to save the environment but also to prevent depletion of valuable resources.

6.6 Understanding and predicting space weather

Stefaan Poedts[1]
[1]KU Leuven, Leuven, Belgium

6.6.1 General overview

In the last few decades, space-based and ground-based solar observations have disclosed that the atmosphere of the Sun is very inhomogeneous, with structures on a wide range of length scales and also very dynamic with activities on a large variety of temporal scales. Some of these solar dynamic events are so energetic that they affect the Earth at a distance of 150 million kilometres. The Sun emits a continuous outflow of solar particles in all directions, called the solar wind. This wind contains transients and interacts with our magnetosphere and the magnetospheres of other planets in the solar system. The whole set of complex effects of the radiation and the plasma stream from the Sun on the Earth and our magnetosphere, our technological systems, our climate, and humankind determines most of the so-called space weather[10]. The broader term 'heliophysics' is used for the study of the interconnectedness of the influence domain of the Sun, that is, the entire solar–heliospheric–planetary system[11].

The activity of the Sun and Sun–Earth interactions are the main drivers of the space weather. It causes substantial socioeconomic damage to human infrastructure both in space and on the Earth. This includes both direct effects on specific industry sectors, such as electric power, spacecraft, and aviation, and indirect effects on dependent infrastructure and services, such as positioning and navigation systems, electric power grids, telecommunication, and oil and gas pipelines. As human society becomes ever more dependent on technological infrastructure that fully or partially relies on the space surrounding us, the impact of space weather events becomes increasingly important.

Over recent years, programs and infrastructure have been built up by national and international agencies and corporations to predict solar activity and to forecast its potential impact on the Earth and its space vicinity. The main science questions are as follows:

- What conditions in the magnetic field emerging into the solar atmosphere lead to eruptive events, such as flares and coronal mass ejections?
- What mechanism or mechanisms cause the sudden acceleration of particles during a solar flare and the gradual solar energetic particle events in coronal mass ejection (CME) shocks?

[10] Definition used in the US National Space Weather Program Strategic Plan [220]: 'The term "Space weather" refers to the conditions on the Sun and in the solar wind, magnetosphere, ionosphere, and thermosphere that can influence the performance and reliability of space-borne and ground-based technological systems and can endanger human life or health.'

[11] The heliosphere is the region surrounding the Sun and the solar system that is filled with the solar magnetic field and the protons and electrons of the solar wind.

- How can we estimate the direction and strength of the particle and radiative fluxes released by these events?
- How is magnetic flux transported from the solar interior to the heliosphere?
- Is it possible to fully understand and accurately predict solar magnetic storms and quantitatively forecast the impact they will have on the Earth?
- How can human society best mitigate effects of extreme space weather events?

6.6.2 Socioeconomic importance

Our Sun is a very dynamic star that influences the Earth in many ways. The continuously changing solar wind as well as eruptions on the solar surface significantly alter the conditions in the whole solar system and more specifically the near-Earth environment (the magnetosphere, ionosphere, and thermosphere). We call these space environmental conditions 'space weather,' and its effects are noticeable in our magnetosphere, the atmosphere, and even at the Earth's surface (see figure 6.47). Because human society is becoming ever more dependent on advanced technologies and infrastructure, space weather phenomena influence our daily lives increasingly by affecting both our space-borne and ground-based technological systems [221–223]. Energetic particles that are ejected from the Sun in the heliosphere or accelerated by shock waves in the solar wind can induce magnetic fields on the Earth, which in turn can drive large currents through power networks that can interfere with the network operation, damage transformers, and

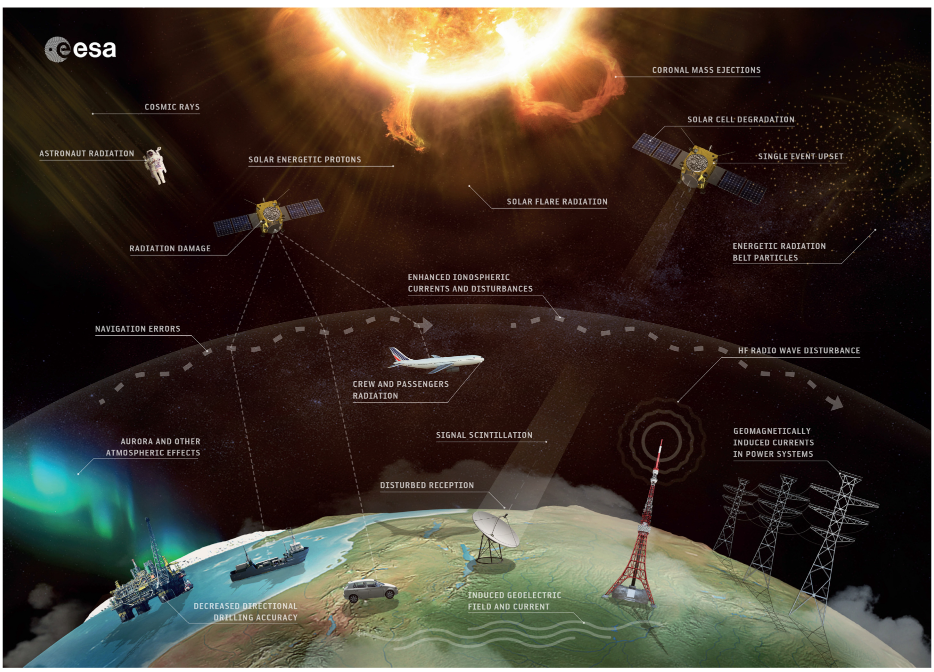


Figure 6.47. Space weather effects. (Source: ESA SSA—Space Weather Segment https://www.esa.int/ESA_Multimedia/Images/2018/01/Space_weather_effects.)

cause electric power loss and damage to oil pipelines. Such high-energy solar particles can also damage satellites and spacecraft in orbit, for instance, by bit swapping in electronics, single-event latch-up (destruction), and wear to solar panels. Astronauts could be killed during extravehicle activities by protons of more than 30 MeV, since these can penetrate spacesuits. Hard particle energy spectra can contain large fluxes of hundreds of MeV–GeV type superenergetic particles, which can reach low Earth orbit satellites and even penetrate the safest areas of spacecraft. The radiation can even endanger airplane crews and frequent airline passengers. Heat expansion of the Earth's thermosphere caused by solar storms can perturb or even change spacecraft orbits and thus disturb and even disrupt telecommunication and navigation systems. Because of this, over the past decade, space weather has been included in national risk assessment plans.

One of the key drivers of space weather are solar eruptions. These include both solar flares and CMEs (see figures 6.48 and 6.49. From a socioeconomic perspective, the CMEs are the main drivers of adverse conditions in the inner heliosphere (e.g., Richardson *et al* [224]). CMEs are large magnetised clouds, containing up to $10^{13} - 10^{16}$ g of plasma, that are ejected into interplanetary space at velocities of typically 450 km s^{-1} (1 620 000 km h^{-1}). CMEs launched in our direction typically arrive at the Earth in about 2–3 days. However, so-called extreme events involve velocities up to 3000 km s^{-1} and more (more than 10 million km h^{-1}) and can reach the Earth in as little as 12 h. While some of the less eruptive CMEs may have minor to no effects on the Earth, the strongest and fastest of them can create severe geomagnetic storms. One of the most severe space weather storms ever reported happened in September 1859 (called the Carrington event), when auroras were spotted at very low latitudes and disruptions of telegraphs were reported.

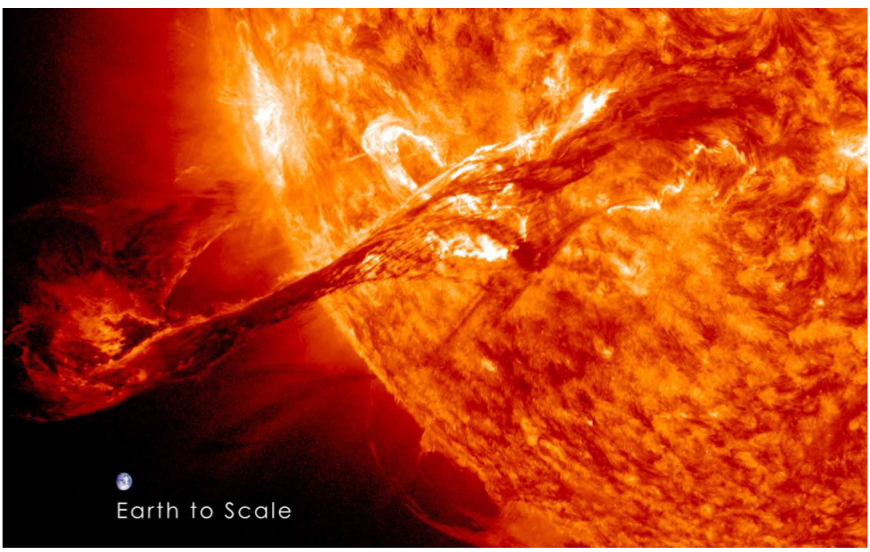


Figure 6.48. Eruption of a solar flare on August 31, 2021. The Earth is projected to scale on the image. (Source: NASA Goddard Space Flight Center.)

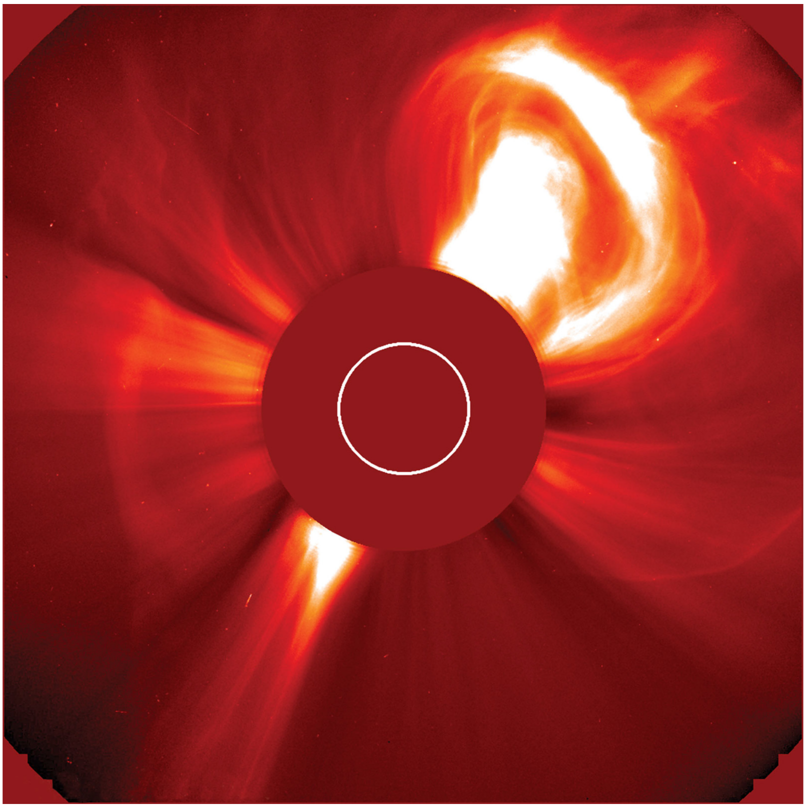


Figure 6.49. Coronagraph image of a CME. The red disk in the centre is the part of the instrument that covers the solar disk (the white circle indicates the solar disk). The false-colour image was taken from the Solar and Heliospheric Observatory (SoHO) spacecraft on December 2, 2002. (Source: SoHO/ESA/NASA.)

The damage was relatively modest because there were no satellites or electric power grids back then. A space weather effects report of the National Research Council of the National Academies in 2008 [225] estimated that the advent of an extreme event, such as the Carrington event, would cost our current society between 1000 billion and 2000 billion (i.e., 1–2 million million) USD. Moreover, it would take 4–10 years to repair all the damage. This is an order of magnitude worse than the harm caused by Hurricane Katrina, which caused 'only' 153 billion USD of destruction [226]. More recently, Eastwood *et al* [227] estimated that the total economic loss from an extreme event would vary from 0.5$tn to 2.7$tn (tn = trillion = 1000 billion, i.e., 10^{12}) based on calculations taking into account the disruption to the global supply chain. Cannon [228] claims that extreme events like the Carrington event occur only about once in 250 years with a confidence level of 95% and once in 50 years with a confidence level of 50%. However, in 2012, Riley [229] estimated that there is a 12% probability of such a superstorm occurring within the next decade. In fact, on 23 July of that year (2012) an unusually large and fast CME event occurred, but it missed the Earth, thanks to the rotation of the Sun. It would have hit us if it had occurred 9 days earlier. This most severe storm in 150 years did hit the Stereo-A spacecraft, though. Luckily, such extreme events are rare, but the more common 'normal' space weather events also have a considerable socioeconomic impact. For instance, Schrijver *et al* [230] deduced from insurance claims that geomagnetically induced currents (GICs) may cause losses of the order of 5–10 billion USD per year to the US power grid.

It is not only CMEs that are affecting space weather. Fast solar wind streams, originating from coronal holes, seen as dark areas in extreme ultraviolet and soft x-ray images of the solar corona, can also cause significant socioeconomic damage. One example of this is the failure of the Anik E2 spacecraft, caused by a fast solar wind stream that swept past the Earth in January 1994. The total damage is estimated to have cost the Telesat satellite communication company around 50–70 million USD in recovery costs and lost business. The recovery of its operational status took 6 months, during which thousands of people lost their television and data services. Satellite blackouts and temporary disruptions are frequent. Howard *et al* [231] determined that about 4500 spacecraft anomalies have been reported over the 25-year period 1974–99 (and currently there are more than 1000 satellites in orbit). The effects of solar storms, caused by CME impact or interaction with high-speed solar wind streams, in the Earth's ionosphere can distort radio waves used for communication or corrupt GPS signals, with significant impact on our current society. Furthermore, space weather events can increase the radiation dose at the altitudes of flying aircraft and, in general, the exposition to radiation of in-orbit astronauts. Therefore, in addition to limiting the socioeconomic impact of space weather events, their mitigation also implies protection of human life at large.

6.6.3 Current state of the art

Given the enormous socioeconomic impact involved, reliable predictions of space weather and its effects on technological systems, human life, and health are extremely important. Genuine predictions and forecasts require a deeper insight into the underlying space weather physics, the mechanisms behind the different phenomena and their effects. Observations are much needed but sometimes limited and/or difficult to interpret, for example, due to projection effects or complexity. Moreover, some things cannot be observed, such as the important coronal magnetic field. Numerical simulation models can provide complementary information, facts that cannot be observed directly, such as the magnetic field topology, the internal density structure of a CME, and the local velocity in and around a CME.

Thus, when trying to assess the daily space weather and its possible dangers, two main types of data are used: observational data obtained from spacecraft and synthetic data obtained from running numerical simulation models. One can compare the situation with the weather conditions on the Earth, except that the amount of observational data to work with is much more limited. There are only a few space weather satellites gathering *in situ* data, compared to the thousands of weather stations in Europe alone [232]. Moreover, one of the main drawbacks of the current observations is that the physical parameters are measured in the line of sight of the spacecraft, which are positioned in only a few different locations in the heliosphere, and until recently they were all in the equatorial plane. This gives rise to projection effects, for example, on the limb of the solar disk or problems to determine the three-dimensional structure of the ejected CME. On top of that, the *in situ* observations that are available are so close to the Earth that they do not give us enough lead time to adapt our satellite and Earth-based systems to an incoming space weather storm. It is for these reasons that we resort to both empirical and physics-based models when it comes to forecasting space weather.

6.6.4 Achievements

Solar drivers of space weather. The Sun emits a continuous stream of charged particles, known as the solar wind. It is inhomogeneous and bimodal, containing slow wind and high-speed solar wind streams. As a result, the interaction of the solar wind with our magnetosphere is not steady and causes the magnetic field of the Earth to vary in time, which in turn drives geomagnetic activity. In fact, the Sun is the main source of space weather events on the Earth. In particular, the main solar drivers are solar flares, CMEs, and solar energetic particle (SEP) events. Solar flares are enormous explosions that are observed as bright areas on the Sun in x-rays and are also visible in optical wavelengths and cause bursts of noise in radio wavelengths. Flares release magnetic energy previously stored the solar atmosphere. The electromagnetic emission produced during flares travels at the speed of light and reaches the Earth after about 8 min. CMEs are violent outbursts of plasma from the Sun's outer atmosphere (the corona). They propagate at thousands of kilometres per second and typically carry a billion tons of material into the heliosphere. When sampled *in situ* by a spacecraft, they are termed Interplanetary CMEs (ICMEs). Apart from the plasma (mostly protons and electrons), CMEs also contain powerful magnetic fields which turn out to play a key role when CMEs interact with our magnetosphere. Unlike solar flares, CMEs are not particularly bright, and they are much slower and typically need 2–4 days to travel to the Earth. The fastest CMEs, however, can reach the Earth in less than a day. In SEP events, large numbers of high-energy charged particles, predominantly protons and electrons, are released. They can be caused by magnetic reconnection-driven processes during solar flares resulting in impulsive SEPs, but also at CME-driven shock waves that produce large gradual SEP events.

Modeling of the solar wind, CMEs, and SEPs. A lot of data are available nowadays, including full-Sun observations from below the solar surface into the heliosphere at multiple wavelengths and far-side and stereoscopic imaging by the instruments aboard the STEREO, SDO, and SoHO satellites. The growing open-access event archival databases facilitate their use for data-driven and physics-based simulation models of the variable solar wind, solar flares, and CMEs. These enable us to explore and quantify the conditions leading to solar storms and their initial development with increasing detail. Also within the heliosphere we now have observations of propagating perturbations traveling from the Sun to the Earth and beyond, including continuous (long-term) *in situ* monitoring of the solar wind and SEP properties at the L1 point, about one million miles upwind from Earth along the Sun–Earth line, and multiscale measurements in the near-Earth solar wind. Such observations provide vital information on the internal structure of CMEs and the CME-driven shock fronts and help us to understand the dynamics of our magnetosphere in response to the variable solar wind conditions and CME impacts.

These observations are crucial for mathematically modeling the phenomena. The data are not only key to validation of the models; state-of-the-art models are data-driven, that is, they use the available data as input, for instance as boundary or initial conditions. Examples of such data-driven, operational solar wind and CME propagation and evolution models are ENLIL [233] and EUHFORIA [234] (see figures 6.50 and 6.51). Both these models use magnetograms of the solar photosphere as boundary conditions for the solar coronal model and various other observational data to determine the launch parameters of the CMEs, such as the initial velocity, spread angle, location, and magnetic flux. These models are made available to the wider community via the NASA Community Coordinated Modeling Center (CCMC, http://ccmc.gsfc.nasa.gov) or the ESA Virtual Space

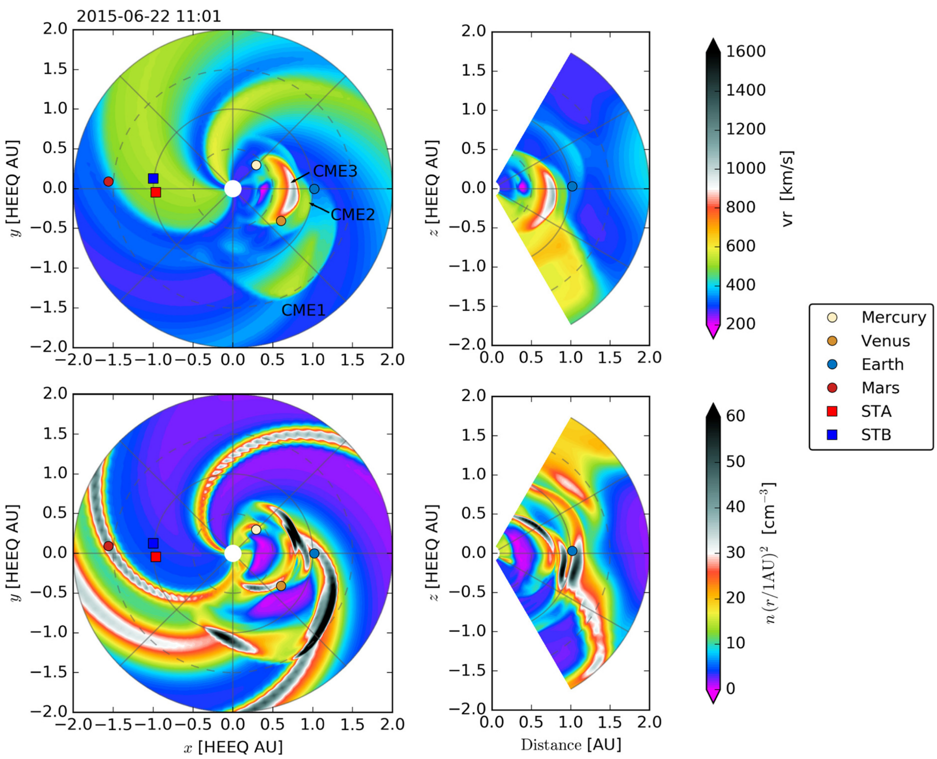


Figure 6.50. Snapshot of an EUHFORIA simulation of an event in June 2015. The snapshot was taken at 03:03 UT on June 21, 2015. The top row shows the radial speed, while the bottom row shows the scaled number density. The left panels depict the solution in the heliographic equatorial plane, while the right panels show the meridional plane that includes the Earth, indicated by the blue dot. Reproduced from [234] CC BY 4.0.

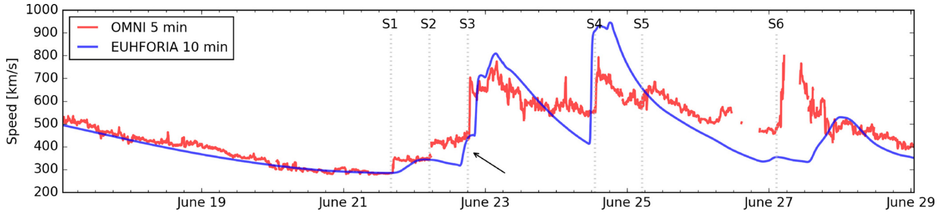


Figure 6.51. Radial speed at the position of the Earth as a function of time. The blue curve shows the data from the EUHFORIA simulation also shown in figure 6.50, saved at 10 min cadence, while the red curve shows OMNI 5 min data. (OMNI is an hourly resolution multisource data set of the near-Earth solar wind's magnetic field and plasma parameters.) Reproduced from [234] CC BY 4.0.

Weather Modelling Centre (VSWMC) [235]. They are used on a daily basis to forecast the arrival of the CIRs and CME shocks and magnetic clouds at the Earth (and other planets) with relative success. They are also continuously adjusted and improved, as they are of course not perfect. Their weaknesses include the use of very simplified (semiempirical) background solar wind models as well as the use of oversimplified CME models that do not take the structure of the magnetic field within the CME itself into account or do so only marginally. They also do not describe the CME onset self-consistently nor its early propagation as the CMEs are introduced only at 0.1 au[12]. These models also do not provide any information about the SEP emission and transport properties generated by solar flares and the CME leading shock fronts and offer no predictions of their geoeffectiveness, the impact they have on the Earth, because they are not coupled with magnetospheric/ionospheric models or with effects models. However, numerical simulation models for particle acceleration at shocks do exist. The Coronal Shock Acceleration simulation model [236], for instance, traces energetic ions and self-consistently determined power spectra of Alfvénic fluctuations that scatter them in prescribed large-scale fields upstream of a shock (including global heliospheric field configurations). Some numerical simplifications made in describing the wave–particle interactions allow the model to be run over global time scales. The SOLar Particle Acceleration in Coronal Shocks model is more recent and uses a physically accurate description of the microphysics involved. However, it is presently limited to local simulation volumes around the shock [237]. The test particle Monte Carlo simulation model called DownStream Propagation with Magnetohydrodynamics models the downstream side of the shock, and the Shock-and-Particle model [238] solves a focused transport equation assuming a constant solar wind flow with a Parker spiral magnetic field. The novel PARADISE model [239] simulates energetic particle distributions in the solar wind provided by EUHFORIA [234] using a quasilinear approach to capture the interaction between solar wind turbulence and energetic particles. An illustration of the possibilities is shown in figure 6.52.

Geoeffects. The variable solar wind conditions causes changes in our magnetic shield, the magnetosphere, the area of space around the Earth that is controlled by

[12] 1 astronomical unit is 149 597 871 kilometres, more or less the average Sun–Earth distance.

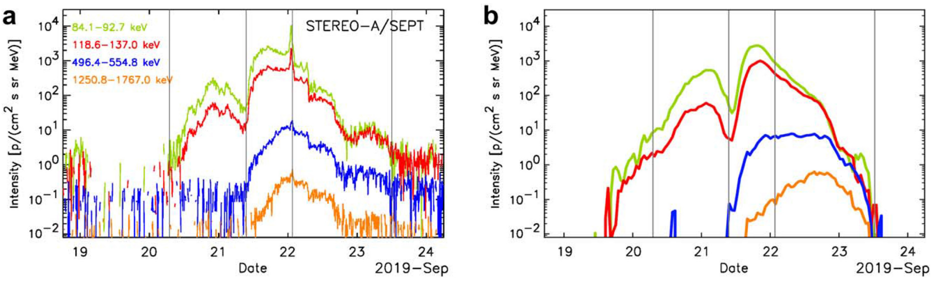


Figure 6.52. PARADISE result: Observed (left) and simulated (right) omnidirectional ion intensities at STEREO-A due to a solar wind stream interaction region (SIR). Vertical lines indicate the onset time of the SIR event (20 September 09:00 UT), the stream interface (21 September 09:30 UT), the developing a reverse shock (22 September 01:35 UT), and the stop time of the SIR event (23 September 12:00 UT). Reproduced from [240], copyright IOP Publishing Ltd. All rights reserved.

its magnetic field. The Earth's magnetosphere is blasted by the solar wind, compressing its sunward side to a distance of only 6–10 times the Earth's radius and creating the bow shock, a supersonic shock wave on the day side. This bow shock heats and slows down the majority of the solar wind particles and causes them to detour around the Earth in the magnetosheath. The night-side magnetosphere is extended by the solar wind to possibly 1000 times the Earth's radius, known as the magnetotail. The geomagnetic response to transients in the solar wind (CIRs, CMEs, magnetic clouds, etc) is expressed in terms of indices such as Kp, Dst, and magnetopause stand-off distance, and these are often used as metric in forecasts of space weather impact on the Earth.

The Kp index is a global geomagnetic activity index based on 3-hourly measurements of magnetometers spread over the whole planet. It consists of 10 values ranging from 0 to 9, where a Kp of 0 indicates quiet conditions, and a Kp of 9 means extreme storm conditions. The disturbance storm time (Dst) index gives information about the strength of the ring current around the Earth caused by solar protons and electrons. The magnetic field induced by this ring current is directly opposite the Earth's magnetic field. Hence, when the difference between solar electrons and protons becomes higher, the Earth's magnetic field becomes weaker. During solar storms, the Dst value becomes negative, meaning that the Earth's magnetic field is weakened.

The Earth's magnetopause separates the solar wind from the Earth's magnetosphere. Its location changes due to different solar wind conditions. Its location is determined by the balance between the pressure of the magnetic field of the Earth and the dynamic pressure of the solar wind. When the solar wind pressure increases, the magnetopause moves inward, and the reverse occurs when it decreases.

Modeling of geoeffects. The observational coverage of the ionosphere–thermosphere–mesosphere domain is growing, which enables near-real-time maps of the total electron content and critical frequency maps for radio communication. Our insight in the contribution of the plasmasphere to the ionospheric total electron content is improving, and so is our understanding of subauroral effects and the troposphere–ionosphere coupling as well as the storage and release of solar wind momentum in the thermosphere.

Global magnetosphere models aim to simulate the detailed spatial and temporal response of the magnetosphere–ionosphere system under solar wind forcing. GUMICS-4 [241] is a coupled magnetosphere–ionosphere simulation model including a global magnetohydrodynamics (MHD) magnetosphere model and an electrostatic ionosphere model. It is capable of predicting the magnetopause distance, interplanetary magnetic field penetration in the magnetotail, ionospheric field-aligned current pattern, and so on. At Imperial College London, a model is being developed based on the global MHD code Gorgon, a three-dimensional (3D) resistive MHD code originally developed to simulate laboratory plasmas [242]. Other models focus on specific parts of the magnetospheric system, such as the 3D model of the plasmasphere coupled with the ionosphere developed by Pierrard *et al* [243] and the Salammbô code, the 3D electron and proton radiation belt model that gives the instantaneous particle flux inside the radiation belts, taking into account wave–particle interactions as well as atmosphere–particle and plasmasphere–particle interactions and even radial diffusion [244]. But we still do not fully understand the acceleration mechanisms of energetic electrons and protons within the radiation belts, and quantitative prediction models are needed.

Simple empirical models and more complicated neural network models exist to derive the geoeffectiveness parameters. They are usually based on L1 data, which means they provide forecasts of the immediately expected parameter values, so-called nowcasts. The current state-of-the-art in Europe, in terms of geomagnetic activity at ground level and in GIC estimation in power grids, is also real-time monitoring. In the UK the MAGIC (Monitoring and Analysis of GIC) power system tool, developed by UKRI-BGS in association with National Grid, represents this state of the art. But real-time limits the reaction time and mitigation options available to the power industry.

6.6.5 Most urgent needs

As was mentioned earlier, there exist some basic operational models for forecasting space weather and its effects. However, many of them are (semi-)empirical and need to be upgraded to physics-based models in order to improve the forecasts, and predict the parameters a few days ahead to enable mitigation.

The operational solar wind models are 3D and data-driven (based on magnetograms of the solar surface, the photosphere), and they are still very much simplified in order to speed up the numerical simulations. The mentioned operational models ENLIL and EUHFORIA both are based on a semiempirical coronal model and provide a steady background solar wind corotating with the Sun starting only at 0.1 au, that is, beyond the 'sonic' point where the solar wind becomes superfast, faster than the characteristic speeds in a magnetised plasma, so the boundary conditions to implement at the inlet boundary (0.1 au) are simple. We need an operational full-MHD coronal model, containing the most important physical effects, such as radiation, thermal conduction, and heating/acceleration. Such models exists, but they require way too much CPU power and wall clock time to be used in operational settings.

We also need more advanced 'flux rope' models for CMEs that are capable of predicting the evolution of the internal magnetic structure of the ICMEs and, in particular, the component of the magnetic field perpendicular to the equatorial plane, as this is key for the severity of the impact of the CMEs on the Earth. The internal magnetic structure of the CMEs evolves due to its interaction with the ambient solar wind, which is also magnetized. This yields deformations and erosion of both the plasma and the magnetic flux in the magnetic cloud. These processes need to be understood very well in order to be able to predict the magnetic structure of the ICME upon impact. In fact, this leads to a more fundamental problem and challenge: We need to better understand the magnetic reconnection process because this generic plasma physics phenomenon is playing a key role in the onset of flares and CMEs (forming the flux rope and cutting it loose from the Sun), the already mentioned deformation and erosion of ICMEs, the interaction of CMEs with the magnetospheric bow shock, and so on. A full quantitative understanding of magnetic reconnection is thus absolutely necessary to understand most of the space weather phenomena. This requires the development of combined full-scale fluid and kinetic models for it.

Another major, related modeling challenge is to understand why certain solar events insert SEPs into the heliosphere and how such SEPs propagate through the heliosphere. We mentioned some advanced models for SEPs before, but these still contain ad hoc terms to take into account the turbulence before the shock, the particle acceleration in the shock, and cross-field diffusion. In order to model impulsive SEP events properly, one would need a self-consistent CME launch model. Also genuine gradual SEP event modeling should include the early evolution phase of the CMEs in the low corona. Such models exist to some extent, but again, they require way too much CPU time to be useful in an operational environment.

In short, we need to take into account more physics in the space weather models. This is because we need real quantitative predictions, preferably a few days ahead of time so that mitigation is possible. Many current space weather models are empirical or semiempirical and facilitate only nowcasts. We urgently need to develop more physics-based models that enable realistic simulations to the extent that they can predict events a few days ahead so that mitigation is possible. Of course, such models are data-driven, and we thus also need more and continuous *in situ* and remote sensing data.

6.6.6 Coupling different simulation models

Because of the huge temporal and spatial extent of space weather phenomena, it is almost impossible to model the entire phenomena from the start (e.g. before eruption) to the end (e.g. a possible space weather storm arriving at the Earth). For this reason, over the past decade, scientists have developed a large variety of models for each of the parts of the larger puzzle. Ranging from close to the Sun at the solar surface, we can model from the preeruption to the propagation of solar wind and CMEs through the solar system all the way to the Earth, where we can model the magnetosphere and even effects on the power grid. Most of these models solve a set of mathematical equations, taking into account the physical and/or chemical processes behind the phenomena of interest. Some of the model inputs are based on observations, for example, the CME structure, but they contain both observational and fitting errors. Numerical simulations based on these models require high-performance computer clusters. But after appropriate validation, some of these models have a predictive value, so they can be used for forecasting, for instance, the arrival of a CME shock at the Earth or the radiation to be expected at the location of a satellite, enabling, in some cases, mitigation of its destructive effects. As each model tries to gain deeper insight into the physical mechanisms that are at hand, they usually focus on a specific area, for example, the solar surface, the Earth's magnetosphere, and so on. However, the output of one numerical model can be used as input for other models, thus creating a chain of models [235]. Thus, we are able to model the propagation of the solar wind and CMEs from the solar surface all the way to the Earth and even specific effects on the Earth, such as power grid effects. The different complicated numerical models usually have a rather complicated installation procedure, and operating these models requires experience. Moreover, the efficient use of these models often requires a considerable amount of computer power (CPU time) and computing memory. Such model runs also produce an enormous amount of data as output, which needs to be stored, interpreted, analysed, and visualised. A preliminary example of the possibilities that the coupling of models creates is given in figure 6.53 where the current nowcasts (a few hours ahead) are compared with forecasts (3 days ahead) obtained by using synthetic (simulation) data at L1 instead of measurements.

Because of the need for this chain of models, scientists started to develop integrated space weather model frameworks and infrastructures in which these models can easily be run and also be coupled. A good example is the CCMC (see [247]). This is a multiagency (NASA and the National Science Foundation) initiative that 'enables, supports and performs the research and development for next-generation space science and space weather models'[245]. Another good example of such an integrated framework is the University of Michigan Center for Space Environment Modeling (CSEM) (see [246]). CSEM also develops physics-based, high-performance computer models and applies them to predict the space weather and its effects. These frameworks provide a kind of standard environment for the computer models and at the same time serve as model and data repositories. They thus enable model simulation runs and some even facilitate the coupling of different submodels that are integrated in the framework. This way the framework is able to support space weather forecasters and researchers and even space science education. Also the VSWMC developed by ESA provides a common framework for running space weather models and easily couple them together to create more complex model chains [235].

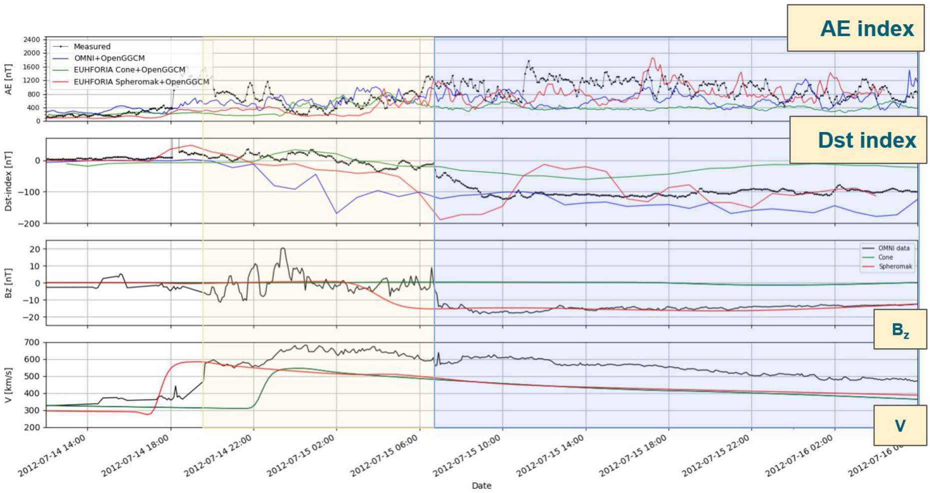


Figure 6.53. Preliminary result from coupling EUHFORIA to the magnetospheric model OpenGGCM. Reprinted by permission from Springer Nature, [247], copyright (2008). The standard nowcast, using measurements at L1, are compared with forecasts (3 days ahead) using synthetic L1 data from EUHFORIA. When a spheromak (magnetized) CME model is used, the forecast can stand the comparison with the nowcast. From top to bottom: comparison of AE index measurements to nowcast (OMNI-OpenGGCM) and forecasts based on EUHFORIA with a cone CME and a spheromak CMEs; the same for the Dst index; the measured B_z component compared to the predictions; and the same comparison for the radial velocity. (Source: Maharana *et al* (to be submitted).)

6.6.7 Challenges

6.6.7.1 Scientific challenges

The goal in space weather–related scientific research is to develop robust methods and models that enable the prediction of space weather events and their geoeffects with high accuracy. The aim is to do better than nowcasts and provide the predictions at least 12 h in advance, preferably even 2–3 days in advance, in order to enable to mitigate the impact on human infrastructure and society as well as to protect human health in flight or in orbit.

The global space weather road map for 2015–25 [222] commissioned by COSPAR and the International Living With a Star program, recognises the progress that has been made regarding both ground-based and space-based observations of the Sun–Earth system. However, a deeper understanding of the complex space weather events involving a variety of physical and chemical effects on many different length scales requires a multidisciplinary and internationally coordinated approach. The roadmap advocates physics-based and data-driven models and prioritises the scientific focus areas and research infrastructure needed to deepen our understanding of space weather and how it affects our daily life. The identified highest-priority research areas include the quantification of the active-region magnetic structure, which is necessary to model and predict nascent coronal mass ejections; the quantification of the solar wind–magnetosphere–ionosphere coupling dynamics;

knowledge of the global coronal field required to drive solar wind models; and understanding and modeling of the radiation belts and solar energetic particles. Moreover, these recommendations are expanded into pathways with focus questions and urgent actions, for instance, to maintain crucial infrastructure and expand it where necessary. The focus should be on observation-based modeling throughout the Sun–Earth system, which should yield prediction models of the internal magnetic structure of coronal mass ejections that enable forecasts more than 12 h ahead of their impact at the Earth. Also the geoeffectiveness of the CME impacts needs to be better understood and how they induce intense geomagnetically induced currents that affect our power and transport infrastructures and radiation storms that cause aging and malfunctions of space-based and sometimes even ground-based assets. The combined effect of radiation and current flow results in ionospheric disturbances that affect navigation and telecommunication systems, cause satellite drag, and lead to ageing of satellites.

The more recent European Space Weather Assessment and Consolidation Committee (ESWACC) report [248] was commissioned by the European Space Science Committee of the European Science Foundation. It first discusses the ongoing European space weather activities and problems and then gives recommendations for better consolidated efforts. It identifies six areas where European-wide coordination is required. These include 'Enabling critical science to improve our scientific understanding of space weather', the 'Development *and coupling* of advanced models by applying a system-science approach which utilises *physics-based* modeling', the 'Assessment of risks at National, Regional and European levels', the 'Consolidation of European User Requirements', 'Support to R2O (and O2R)', and 'Define and implement an operational network for future space weather observations'.

It is extremely difficult to correctly predict the component of the magnetic field of a ICME upon impact at the Earth's bow shock, especially when the CME hits the Earth not fully but gives it only a glancing blow [249, 250]. The fact that the Earth is about 150 million kilometres away from the Sun makes it such that a minor error in the initial position parameters may result in the CME 'missing' the Earth. To mitigate such errors on the input parameters, so-called ensemble simulations are performed where a few dozen of combinations of input parameters are considered (based on the uncertainty of the observations), yielding a prediction windows for, for example, the shock arrival time, that is, a window in which 75% of the ensemble CME shocks arrived.

Extreme space weather events are a particular challenge, as they occur only every 100 or 200 years and moreover do not follow the solar cycle, that is, they can occur anytime, including during solar minimum [251]. Moreover, it has been shown that CME-CME interactions can turn two average CMEs into a very geoeffective combination under the right circumstances [250]. Hence, the potential geoeffectiveness of such CME–CME interactions needs to be investigated more, as they occur very frequently.

6.6.7.2 The data avalanche challenge

The solar and space environment (and the broader heliophysics) research community is experiencing a rapid and radical transformation with an enormous increases in data rates and archive volumes. There are urgent needs for theoretical support for the interpretation of the intricate, nonlinear systems that couple the Sun's deep interior all the way to its atmosphere and further out to the planetary climate systems. In the subdiscipline of solar physics, for example, the currently operative Solar Dynamics Observatory (launched in 2010) sends down over 2 terabytes of data per day. This constitutes a hundredfold increase over the TRACE mission and a thousandfold increase of the SOHO/EIT instrument, both of which are just over 25 years old. This flood of data cannot be routinely inspected and analysed by the scientific community in the traditional ways of personal inspection and detailed analysis of much of the data. Instead, the instrument teams are guiding the community towards automated feature recognition, pattern finding, and knowledge base–supported guidance to preselected subsets of the full data archive that are deemed to be the most helpful as we seek to further our understanding of solar activity and the space weather phenomena that it drives.

On the theoretical/numerical front, we are increasingly aware of the fact that the solar interior, the solar atmosphere, and the heliosphere form a nonlinear coupled system with multiple complex feedback pathways. The interpretation of the observational data has moved past the intuitive or elemental model phase and urgently requires advanced numerical models that are able to simulate the intricate coupled processes in the system.

6.6.7.3 The research coordination challenge

Developing such advanced physics-based numerical simulation models for the intricate 'big problems' in solar physics and in heliophysics in general requires a different way to manage, perform, and fund scientific research. The current practice resulted in small research groups with unpredictably fluctuating financial resources and few long-term, fully supported scientists, postdoctoral researchers, and PhD students. These have revealed the potential value of such models, but we have now reached the phase where larger, collaborative, and more coordinated efforts are required. The obvious analogy is the similar experience in climate research with the development of general circulation models (GCMs). In the 1990s, the GCMs became so complex, as they were incorporating ever more coupled processes, that individual research groups were forced to collaborate on large community-based models to make progress. The basis of a similar effort in solar and space weather physics is currently being created by the ESA VSWMC. However, the VSWMC will integrate and couple only currently existing models, while these also need to be advanced to a larger scale that enables us to address the big problems of solar and space weather physics. This is a very important key competence, from both strategic (independence, military, etc) and socioeconomic points of view, of the EU research community that needs to be developed and maintained.

The needed coordinated development of complex, efficient, high-resolution, and high-fidelity numerical codes is very different from, for example, having an environment in which existing codes are available to the community (such as at the CCMC in the US and ESA's VSWMC) or in which codes are strung together to cross disciplinary interfaces (such as at the Center for Space Environment Modeling in the US), and something on a substantially larger scale than the International Solar-Terrestrial Physics Geospace Model [252] is needed. In order to develop powerful 'missions to the virtual reality' we will need to treat them as space missions, starting with science definition teams, followed by competitive selection of projects, led from a concept–study report phase into a detailed construction phase with well-defined interfaces between elements and subsequently operated by a staff of experts who aid a user community in operating the codes and in archiving, retrieving, and interpreting the results.

6.6.7.4 The evolving space weather landscape

The ESWACC report [248] also mentions other issues requiring attention, such as the need for 'continuous elaboration of the analysis including assessment of space weather risks on European infrastructure and understanding of the user needs' and a better coordination of the 'scattered' EU space weather funding ('± 60 Mâ, over the last 10 years') and the preoperational space weather services currently developed by ESA in the optional Space Situational Awareness Programme. Recently, a bottom-up initiative arose within the European space weather community. A group of researchers started to organise itself in a reaction to the changing space weather landscape in Europe and the concern to sustain and further develop the research and user community and the space weather assets (the journal, the yearly European Space Weather Week conference, the medals) and thus to somehow continue to develop the successful efforts made thus far [253].

6.6.8 Opportunities

6.6.8.1 Increasing high-performance computing power

High-performance computing is a game-changing technology in this field (and in many other fields as well). As a matter of fact, with the rapid increase in computational power, a new opportunity is arising: Numerical experiments and simulations now provide access to a virtual world that has the potential of being as revealing about the world around us as direct observations. This is particularly true for the many intrinsically nonlinear processes where numerical experiments can reveal instabilities and bifurcations to guide us to the correct parameter range that applies to the heliophysical processes under study. A few such examples are already found among the most-cited papers in solar and space weather physics[13] (apart from the mission and instrument papers): a model for CME initiation [254], the study of radiative transfer in near-surface convection (also for abundance determinations) [255], and a study of particle acceleration at the Sun and in the heliosphere [256].

[13] Determined from a search of refereed publications with ADS, containing any of the following terms in the abstract: 'solar atmosphere', 'solar corona', 'solar photosphere', 'solar chromosphere', 'solar transition region', 'solar region', 'solar flare', 'space weather', or 'coronal mass ejection'.

6.6.8.2 Funding programs

ESA started preparing for space weather in 2001 [257], but the Space Weather (SWE) Portal was opened only in 2012. It has been expanded ever since, and the first SWE Service elements transited to operations in 2020 in the Space Situational Awareness (SSA) program. ESA's Space Weather Network is currently managed in a preoperational framework and is coordinated by the Space Weather Coordination Centre, located in Brussels, Belgium. It organises a large network consisting of more than 40 teams from organisations across Europe who are spread over five Expert Service Centres. These are collaborating to provide tailored products and services for Space Weather Network customers. In the meantime, ESA continuous to create opportunities via different technology programs and the SSA program.

ESA's new Space Safety Programme (S2P) replaces and continues the SSA program and adds space mission elements to it. It will test the Space Weather Service Network's operational capacity to support ESA's space operations and mission design. This includes nowcasting and forecasting all aspects of space weather in the heliosphere but also in the magnetosphere, ionosphere, and atmosphere. The S2P also includes missions such as the Lagrange mission, the first ever operational space weather mission and the first mission to L5, that is, outside of the Earth–Sun line. The mission is aimed at space weather monitoring and has been developed in close collaboration with NOAA and NASA; its launch is foreseen in 2027. In fact, the Lagrange mission envisions launching two spacecraft at Lagrangian points L1 and L5, which will enable scientists to research solar flares and CMEs and improve forecasts of the solar wind at 1 au. S2P has other components, such as the in-orbit servicing/removal mission (ADRIOS), which will accomplish the first ever removal of a piece of space debris; the HERA mission, which will aim to develop planetary defence technologies (to be launched in 2024); and the CREAM mission, involving a variety of activities aimed at automatic collisions avoiding.

The role of space weather has also become ever more important for EU/EC research programs. The results of the many different research programs are available on CORDIS, the Community Research and Development Information Service [258]. In June 2021, a search on 'Space Weather' in CORDIS yielded 369 hits in the Horizon 2020 program for all collections together. Limiting the selection to projects yielded only 50 hits. The first two 'space weather' projects were funded already in Framework Programs 4 (FP4). In both FP5 and FP6 there were six 'space weather' projects funded (and 14 hits over all collections in FP6), and in FP7, 82 projects were funded (with 215 'space weather' hits over all collections).

Horizon Europe, the EU's key funding program for research and innovation, has a budget of ε95.5 billion for the period 2021–27. However, as of June 2021 there was only one forthcoming 'Space Weather' call: the HORIZON Research and Innovation Action HORIZON-CL4–2022-SPACE-01–62[14]. Clearly, other, more general calls on AI and ML, for instance, also include opportunities for space weather research.

[14] Searching for 'space weather' on https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-search

It is good to notice that the advice from the mentioned ESWACC report [248] on the ESA–EU coordination is being implemented. As a matter of fact, the last Horizon 2020 space weather calls insisted on using existing EU and ESA infrastructures as much as possible.

Of course, additional funding for space weather activities is provided by individual European states. However, this funding source is very fragmented and localised, often focusing on the local space industry. As was mentioned earlier, the complexity of space weather requires coordinated scientific research and harmonised provision of services with the ESA–EU coordination and the ESA Space Weather Service Network as good examples, respectively.

Acknowledgments

UO has been supported by research focus point Earth and Environmental Systems of the University of Potsdam. UO and AA acknowledge Co-PREPARE of DAAD (DIP Project No. 57553291). SS acknowledges support from the DFG research training group 'Natural Hazards and Risks in a Changing World' (Grant No. GRK 2043/1). DE acknowledges TÜBİTAK (Grant No. 118C236). We thank Georg Veh for helping with visualization.

References

[1] Boers N, Goswami B, Rheinwalt A, Bookhagen B, Hoskins B and Kurths J 2019 Complex networks reveal global pattern of extreme-rainfall teleconnections *Nature* **566** 373–7

[2] Steffen W, Sanderson R A, Tyson P D, Jäger J, Matson P A, Moore B, Oldfield F, Richardson K, Schellnhuber H-J and Turner B L 2006 *Global Change and the Earth System: A Planet Under Pressure* (Berlin: Springer Science & Business Media)

[3] Mann M E, Steinman B A, Brouillette D J and Miller S K 2021 Multidecadal climate oscillations during the past millennium driven by volcanic forcing *Science* **371** 1014–9

[4] Ge Y, Jin Y, Stein A, Chen Y, Wang J, Wang J, Cheng Q, Bai H, Liu M and Atkinson P M 2019 Principles and methods of scaling geospatial Earth science data *Earth Sci. Rev.* **197** 102897

[5] Fan J, Meng J, Ludescher J, Chen X, Ashkenazy Y, Kurths J, Havlin S and Schellnhuber H J 2020 Statistical physics approaches to the complex Earth system *Phys. Rep.* **896** 1–84

[6] Bradley R S 2015 *Paleoclimatology: Reconstructing Climates of the Quaternary* (New York: Academic)

[7] Muller R A and MacDonald G J 2002 *Ice Ages and Astronomical Causes: Data, Spectral Analysis and Mechanisms* (Berlin: Springer Science & Business Media)

[8] Poincaré H 1890 Sur le problème des trois corps et les équations de la dynamique *Acta Math.* **13** A3–270

[9] Nicolis G and Nicolis C 2012 *Foundations of Complex Systems: Emergence, Information and Predicition* (Singapore: World Scientific)

[10] Kantz H and Schreiber T 2004 *Nonlinear Time Series Analysis* (Cambridge: Cambridge University Press) vol 7

[11] Eckmann J 1987 Recurrence plots of dynamical systems *Europhys. Lett.* **5** 973–7

[12] Marwan N, Romano M C, Thiel M and Kurths J 2007 Recurrence plots for the analysis of complex systems *Phys. Rep.* **438** 237–329

[13] Pikovskij A, Rosenblum M and Kurths J 2007 *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge: Cambridge University Press)
[14] Torrence C and Compo G P 1998 A practical guide to wavelet analysis *Bull. Am. Meteorol. Soc.* **79** 61–78
[15] Newman M E J 2010 *Networks: An Introduction* (Oxford: Oxford University Press)
[16] Jafarimanesh A, Mignan A and Danciu L 2018 Origin of the power-law exponent in the landslide frequency-size distribution *Nat. Hazards Earth Syst. Sci. Discuss.* 1–28
[17] Gutenberg B and Richter C F 1944 Frequency of earthquakes in California *Bull. Seismol. Soc. Am.* **34** 185–8
[18] Gutenberg B and Richter C F 1956 Earthquake magnitude, intensity, energy, and acceleration: (second paper) *Bull. Seismol. Soc. Am.* **46** 105–45
[19] Frohlich C and Davis S D 1993 Teleseismic b values; or, much ado about 1.0 *J. Geophys. Res.: Solid Earth* **98** 631–44
[20] Blöschl G 1996 *Scale and Scaling in Hydrology* (Techn. Univ., Inst. f. Hydraulik, Gewässerkunde u. Wasserwirtschaft)
[21] Peters-Lidard C D, Pan F and Wood E F 2001 A re-examination of modeled and measured soil moisture spatial variability and its implications for land surface modeling *Adv. Water Res.* **24** 1069–83
[22] Kumar P and Foufoula-Georgiou E 1993 A multicomponent decomposition of spatial rainfall fields: 1. Segregation of large- and small-scale features using wavelet transforms *Water Resour. Res.* **29** 2515–32
[23] Schuster A 1898 On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena *J.Geograph. Res.* **3** 13–41
[24] Eroglu D, McRobie F H, Ozken I, Stemler T, Wyrwoll K-H, Breitenbach S F, Marwan N and Kurths J 2016 See–saw relationship of the Holocene East Asian–Australian summer monsoon *Nat. Commun.* **7** 1–7
[25] Malik N, Marwan N and Kurths J 2010 Spatial structures and directionalities in Monsoonal precipitation over South Asia *Nonlinear Processes Geophys.* **17** 371–81
[26] Pacheco J F, Scholz C H and Sykes L R 1992 Changes in frequency–size relationship from small to large earthquakes *Nature* **355** 71–3
[27] Zhan Z 2017 Gutenberg–Richter law for deep earthquakes revisited: a dual-mechanism hypothesis *Earth Planet. Sci. Lett.* **461** 1–7
[28] Scholz C H 1968 The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes *Bull. Seismol. Soc. Am.* **58** 399–415
[29] Tormann T, Enescu B, Woessner J and Wiemer S 2015 Randomness of megathrust earthquakes implied by rapid stress recovery after the Japan earthquake *Nat. Geosci.* **8** 152–8
[30] Wiemer S and McNutt S R 1997 Variations in the frequency-magnitude distribution with depth in two volcanic areas: Mount St. Helens, Washington, and Mt. Spurr, Alaska *Geophys. Res. Lett.* **24** 189–92
[31] De Geer G 1926 On the solar curve: as dating the ice age, the New York Moraine, and Niagara Falls through the Swedish Timescale *Geogr. Ann.* **8** 253–83
[32] Berger A L 1978 Long-term variations of caloric insolation resulting from the Earth's orbital elements *Quat. Res.* **9** 139–67

[33] Westerhold T, Marwan N, Drury A J, Liebrand D, Agnini C, Anagnostou E, Barnet J S, Bohaty S M, De Vleeschouwer D and Florindo F 2020 An astronomically dated record of Earth's climate and its predictability over the last 66 million years *Science* **369** 1383–7

[34] Burke K D, Williams J W, Chandler M A, Haywood A M, Lunt D J and Otto-Bliesner B L 2018 Pliocene and Eocene provide best analogs for near-future climates *Proc. Natl. Acad. Sci.* **115** 13288–93

[35] Pai D S and Nair R M 2009 Summer monsoon onset over Kerala: new definition and prediction *J. Earth Syst. Sci.* **118** 123–35

[36] Stolbova V, Surovyatkina E, Bookhagen B and Kurths J 2016 Tipping elements of the Indian monsoon: prediction of onset and withdrawal: tipping elements of monsoon *Geophys. Res. Lett.* **43** 3982–90

[37] Ludescher J *et al* 2021 Network-based forecasting of climate phenomena *Proc. Natl Acad. Sci.* **118** e1922872118

[38] Stark C P and Hovius N 2001 The characterization of landslide size distributions *Geophys. Res. Lett.* **28** 1091–4

[39] Malamud B D, Turcotte D L, Guzzetti F and Reichenbach P 2004 Landslides, earthquakes, and erosion *Earth Planet. Sci. Lett.* **229** 45–59

[40] West G B 2017 *Scale: The Universal Laws of Growth, Innovation, Sustainability, and the Pace of Life in Organisms, Cities, Economies, and Companies* (London: Penguin)

[41] Liu Z, Nadim F, Garcia-Aristizabal A, Mignan A, Fleming K and Luna B Q 2015 A three-level framework for multi-risk assessment *Georisk: Assess. Manag. Risk Eng. Syst. Geohazards* **9** 59–74

[42] von Specht S, Ozturk U, Veh G, Cotton F and Korup O 2019 Effects of finite source rupture on landslide triggering: the 2016 *MW* 7.1 Kumamoto earthquake *Solid Earth* **10** 463–86

[43] Ozturk U, Marwan N, Korup O, Saito H, Agarwal A, Grossman M J, Zaiki M and Kurths J 2018 Complex networks for tracking extreme rainfall during typhoons *Chaos* **28** 075301

[44] Saito H, Korup O, Uchida T, Hayashi S and Oguchi T 2014 Rainfall conditions, typhoon frequency, and contemporary landslide erosion in Japan *Geology* **42** 999–1002

[45] Agarwal A, Caesar L, Marwan N, Maheswaran R, Merz B and Kurths J 2019 Network-based identification and characterization of teleconnections on different scales *Sci. Rep.* **9** 8808

[46] Rahmstorf S 2002 Ocean circulation and climate during the past 120,000 years *Nature* **419** 207–14

[47] Caesar L, McCarthy G D, Thornalley D J R, Cahill N and Rahmstorf S 2021 Current Atlantic meridional overturning circulation weakest in last millennium *Nat. Geosci.* **14** 118–20

[48] Crutzen P and Stoermer E 2000 International geosphere biosphere programme (IGBP) *Newsletter* 41

[49] Rai R and Sahu C K 2020 Driven by data or derived through physics? A review of hybrid physics guided machine learning techniques with cyber-physical system (cps) focus *IEEE Access* **8** 71050–73

[50] Karpatne A, Atluri G, Faghmous J H, Steinbach M, Banerjee A, Ganguly A, Shekhar S, Samatova N and Kumar V 2017 Theory-guided data science: a new paradigm for scientific discovery from data *IEEE Trans. Knowl. Data Eng.* **29** 2318–31

[51] Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J and Carvalhais N 2019 Deep learning and process understanding for data-driven Earth system science *Nature* **566** 195–204

[52] Faghmous J H and Kumar V 2014 A big data guide to understanding climate change: the case for theory-guided data science *Big Data* **2** 155–63

[53] Xu T and Valocchi A J 2015 Data-driven methods to improve baseflow prediction of a regional groundwater model *Comput. Geosci.* **85** 124–36

[54] Dietz M *et al* 2022 More than heavy rain turning into fast-flowing water – a landscape perspective on the 2021 Eifel floods *Nat. Hazards Earth Syst. Sci.* **22** 1845–56

[55] Bozzolan E, Holcombe E, Pianosi F and Wagener T 2020 Including informal housing in slope stability analysis–an application to a data-scarce location in the humid tropics *Nat. Hazards Earth Syst. Sci.* **20** 3161–77

[56] Hao Z, Singh V P and Hao F 2018 Compound extremes in hydroclimatology: a review *Water* **10** 718

[57] Zscheischler J, Martius O, Westra S, Bevacqua E, Raymond C, Horton R M, van den Hurk B, AghaKouchak A, Jézéquel A and Mahecha M D 2020 A typology of compound weather and climate events *Nat. Rev. Earth Environ.* **1** 333–47

[58] Raymond C, Horton R M, Zscheischler J, Martius O, AghaKouchak A, Balch J, Bowen S G, Camargo S J, Hess J and Kornhuber K 2020 Understanding and managing connected extreme events *Nat. Clim. Change* **10** 611–21

[59] Hao Z and Singh V P 2020 Compound events under global warming: a dependence perspective *J. Hydrol. Eng.* **25** 03120001

[60] Boers N, Kurths J and Marwan N 2021 Complex systems approaches for Earth system data analysis *J. Phys.: Complex.* **2** 011001

[61] Perez R and Perez M 2015 A fundamental look at energy reserves for the planet *Int. Energy Agency SHC Program. Sol. Updat* (https://iea-shc.org/data/sites/1/publications/2015-11-A-Fundamental-Look-at-Supply-Side-Energy-Reserves-for-the-Planet.pdf)

[62] Solarmap 2021 (https://google.com/search?q=global+solar+energy+distribution&client=firefox-b-e&source=lnms&tbm=isch&sa=X&ved=2ahUKEwjjjYnAzrnyAhWICewKHTdvAnsQ_AUoAnoECAIQBA&biw=1264&bih=573#imgrc=3jhhX0D-D8vxUM)

[63] Shockley W and Queisser H J 1961 Detailed balance limit of efficiency of p-n junction solar cells *J. Appl. Phys.* **32** 510

[64] NREL *Best Research-Cell Efficiency Chart* (https://nrel.gov/pv/cell-efficiency.html) Accessed 13 September 2021.

[65] Stieglitz R and Heinzel V 2012 *Thermische Solarenergie—Grundlagen, Technologie, Anwendungen* (Heidelberg: Springer)

[66] International Technology Roadmap for Photovoltaic (ITRPV) 2020 *Results* 12th edn (https://itrpv.vdma.org/)

[67] Philipps S and Warmuth W 2021 *Photovoltaics Report* (https://ise.fraunhofer.de/en/publications/studies/photovoltaics-report.html) Accessed 27 July 2021

[68] Lee M *et al* 2012 Efficient hybrid solar cells based on meso-superstructured organometal halide perovskites *Science*

[69] Al-Ashouri A *et al* 2020 Monolithic perovskite/silicon tandem solar cell with >29% efficiency by enhanced hole extraction *Science*

[70] Pitz-Paal R 2020 *Concentrating Solar Power in Future Energy* (Amsterdam: Elsevier) 3rd edn pp 413–30

[71] DLR 2021 *German Aerspace Center: Solar Thermal Power Plants, Heat Electricity and Fuels Form Concentrating Solar Power* (https://dlr.de/sf/en/DownloadCount.aspx?raid=533626&docid=13013&rn=adfa4881-0ef9-4df1-b131-7bd11bfb8758)
[72] EASAC 2011 *Concentrating Solar Power: Its Potential Contribution to a Sustainable Energy Future* EASAC Policy Report 16 (European Academies Science Advisory Council EASAC) 2011, Halle (Saale) (https://easac.eu/publications/details/concentrating-solar-power-its-potential-contribution-to-asustainable-energy-future/) Accessed 05 February 2021
[73] IEA 2020 *World Energy Outlook 2020* (Paris: IEA) https://iea.org/reports/world-energy-outlook-2020
[74] Heagel N *et al* 2019 *Terawatt-Scale Photovoltaics: Transform Global Energy* (United States: N.p.) (https://doi.org/10.1126/science.aaw1845)
[75] REN21 2020 *Renewables 2020—Global Status Report* UN Environment Programme, REN21 Secretariat, Paris (https://ren21.net/wp-content/uploads/2019/05/gsr_2020_full_report_en.pdf) Accessed 18 December 2020
[76] Denholm P and Hand M 2011 Grid flexibility and storage required to achieve very high penetration of variable renewable electricity *Energy Policy* (Amsterdam: Elsevier) vol 39 pp 1817–30
[77] IRENA 2020 *Renewable Power Generation Costs in 2019* (Abu Dhabi: International Renewable Energy Agency) https://irena.org/publications/2020/Jun/Renewable-Power-Costs-in-2019)
[78] Solar Energy Technologies Office, U. S. Department of Energy 2017 *The Sunshot 2030 Goals: 3¢ per Kilowatt Hour for PV and 5¢ per Killowatt Hour for Dispatchable CSP. The Sunshot Goals, DOE/EE-1501* (Office of Energy Efficiency & Renewable Energy) (https://energy.gov/sites/prod/files/2020/09/f79/SunShot%202030%20White%20Paper.pdf)
[79] IRENA 2018 *Global Energy Transformation: A Roadmap to 2050* (https://irena.org/publications/2018/Apr/Global-Energy-Transition-A-Roadmap-to-2050)
[80] MCC 2021 *Mercator Research Institute on Global Commons and Climate Change: Remaining Carbon Budget* (https://mcc-berlin.net/forschung/co2-budget.html)
[81] IEA 2021 *Net Zero by 2050, IEA Flagship Report* (https://iea.blob.core.windows.net/assets/beceb956-0dcf-4d73–89fe-1310e3046d68/NetZeroby2050-ARoadmapfortheGlobalEnergySector_CORR.pdf)
[82] Wagner H-J 2018 *Introduction to Wind Energy Systems: Basics, Technology and Operation* (Berlin: Springer) 3rd edn
[83] Vermeiren P, Adriansens W, Moreels J P and Leysen R 1998 The composite Zirfon separator for alkaline water electrolysis *Hydrogen Power: Theoretical and Engineering solutions* (Dordrecht: Springer)
[84] Hall D E 1981 Electrodes for alkaline water electrolysis *J. Electrochem. Soc.* **128** 740
[85] The Engineer 2017 Hot rocks offer solution to grid-scale energy storage *The Engineer* Nov 2017
[86] Santhanam S, Heddrich M P, Riedel M and Friedrich K A 2017 Theoretical and experimental study of reversible solid oxide cell (r-SOC) systems for energy storage *Energy* **141** 202–14
[87] Phase change materials for thermal energy storage | Climate Technology Centre & Network | Tue, 11/08/2016 (ctc-n.org).
[88] advanced-chemistryprize2019.pdf (https://nobelprize.org).

[89] How to prevent short-circuiting in next-gen lithium batteries | MIT News | Massachusetts Institute of Technology
[90] Battery prices have fallen 88 percent over the last decade | Ars Technica
[91] Xu Wangwang and Wang Ying 2019 Recent progress on zinc-ion rechargeable batteries *Nano-Micro Lett.* **11** 90
[92] https://ec.europa.eu/environment/international_issues/relations_china_en.htm
[93] Eddington A 1920 The internal constitution of the stars *Nature* **106** 14
[94] Bethe H A 1939 Energy production in stars *Phys. Rev.* **55** 434
[95] Bigot B 2019 Progress toward ITER's first plasma *Nucl. Fusion* **59** 112001
[96] Progress in ITER Physics Basis 2007 Overview and summary *Nucl. Fusion* **47** S1–S404
[97] Federici G *et al* 2019 Overview of the DEMO staged design approach in Europe *Nucl. Fusion* **59** 066013
[98] Alfvén H 1942 Existence of electromagnetic-hydrodynamic waves *Nature* **150** 405
[99] Huijsmans G T A and Loarte A 2013 Non-linear MHD simulation of ELM energy deposition *Nucl. Fusion* **53** 123023
[100] Giancarli L M *et al* 2020 Overview of recent ITER TBM Program activities *Fusion Eng. Design* **158** 111674
[101] Ibarra A *et al* 2018 The IFMIF-DONES project: preliminary engineering design *Nucl. Fusion* **58** 105002
[102] Classification of Radioactive Waste 2004 *No. GSG-1, General Safety Guide* IAEA.
[103] BP Statistical Review of World Energy 2021 (www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html)
[104] Key World Energy Statistics 2020 International Energy Agency (www.iea.org/reports/key-world-energy-statistics-2020)
[105] IAEA Power Reactor Information System (PRIS) 2021 (pris.iaea.org/PRIS/home.aspx) *IAEA*
[106] Projected Costs of Generating Electricity 2020 (www.iea.org/reports/projected-costs-of-generating-electricity-2020)© *OECD/IEA and OECD/NEA 2020*
[107] Climate Change 2014 *Mitigation of Climate Change—Working Group III Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)* Cambridge University Press (www.ipcc.ch/report/ar5/wg3/)
[108] Poinssot C *et al* 2014 Assessment of the environmental footprint of nuclear energy systems. Comparison between closed and open fuel cycles *Energy* **69** 199–211
[109] Uranium 2018 *Resources, Production and Demand* (OECD) (www.oecd-nea.org/jcms/pl5080/) *NEA No. 7413*, 2018
[110] Implementation of Defence in Depth 2016 *at Nuclear Power Plants—Lessons Learnt from the Fukushima Daiichi Accident* (OECD) (www.oecd-nea.org/jcms/pl14950)
[111] Human and Organizational 2016 *Aspects of Assuring Nuclear Safety* (IAEA).
[112] Storage and Disposal of Radioactive Waste *2016–2021 World Nuclear Association, Registered in England and Wales, Number 01215741*
[113] Status and Trends in Spent 2018 *Fuel and Radioactive Waste Management* IAEA (www.iaea.org/publications/11173/status-and-trends-in-spent-fuel-and-radioactive-waste-management)
[114] Disposal of Radioactive 2011 *Waste—Specific Safety Requirements No. SSR-5* IAEA (www.iaea.org/publications/8420/disposal-of-radioactive-waste)
[115] StakeHolder-based Analysis 2019–2022 of Research for Decommissioning, share-h2020.eu *Research Funded by Euratom, Grant Agreement ID: 847626*

[116] Nifenecker H, Meplan O and David S 2001 *Accelerator Driven Subcritical Reactors* (Bristol: Institute of Physics Publishing)
[117] Aït Abderrahim H *et al* 2020 Partitioning and transmutation contribution of MYRRHA to an EU strategy for HLW management and main achievements of MYRRHA related FP7 and H2020 projects: MYRTE, MARISA, MAXSIMA, SEARCH, MAX, FREYA, ARCAS *EPJ Nucl. Sci. Technol.* **6** 33
[118] Strategic Research Agenda of EURAD 2019–2024 European Joint Programme on Radioactive Waste Management (www.ejp-eurad.eu/sites/default/files/2020-01/2._eurad_sra.pdf)*Research Funded by Euratom, Grant Agreement ID: 847593*
[119] Climate Change and Nuclear Power 2020 (www.iaea.org/publications/14725/climate-change-and-nuclear-power-2020)
[120] Global Warming 2019 *of 1.5 °C—An IPCC Special Report on the Impacts of Global Warming of 1.5 °C above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty* (www.ipcc.ch/sr15/). Intergovernmental Panel on Climate Change.
[121] Net Zero by 2050 2021 *A Roadmap for the Global Energy Sector* (www.iea.org/reports/net-zero-by-2050)
[122] Nuclear Power in a Clean Energy System 2019 International Energy Agency (webstore.iea.org/nuclear-power-in-a-clean-energy-system)
[123] Generation IV International Forum^{SM} (www.gen-4.org)
[124] Sustainable Nuclear Energy Technology Platform (snetp.eu/)
[125] Pakhomov I 2018 BN-600 and BN-800 operating experience *Gen IV International Forum*^{SM} (www.gen4.org/gif/upload/docs/application/pdf/2019-01/gifiv_webinar_pakhomov_19_dec_2018_final.pdf).
[126] Advances in Small Modular 2020 Reactor Technology Developments—A Supplement to: IAEA Advanced Reactors Information System (ARIS) 2020 edn (aris.iaea.org/sites/Publications.html)
[127] Small Modular Reactors 2021 *Challenges and Opportunities* OECD (www.oecd-nea.org/jcms/pl) (57979/small-modular-reactors-challenges-and-opportunities), NEA No.7560
[128] IPCC 2018 Summary for policymakers *Global Warming of 1.5 °C. An IPCC Special Report on the Impacts of Global Warming of 1.5 °C above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty* ed V Masson-Delmotte *et al* (https://ipcc.ch/sr15/chapter/spm/)
[129] United Nations Environment Programme 2019 *Emissions Gap Report 2019* (Nairobi: UNEP)https://unep.org/resources/emissions-gap-report-2019
[130] www.iea.org/data-and-statistics
[131] Hoffmann B 2019 *Air pollution in cities: urban and transport planning determinants and health in cities Integrating Human Health into Urban and Transport Planning* ed M Nieuwenhuijsen and H Khreis (Cham: Springer)
[132] European Environment Agency 2020 *Air Quality in Europe* https://op.europa.eu/en/publication-detail/-/publication/447035cd-344e-11eb-b27b-01aa75ed71a1/language-en
[133] MacKay D J C 2009 *Sustainable Energy—Without the Hot Air* p 231 (http://withouthotair.com/)

[134] European Environment Agency 2020 Monitoring CO2 Emissions From Passenger Cars and Vans in 2018 https://op.europa.eu/en/publication-detail/-/publication/6bac010d-dc45-11ea-adf7-01aa75ed71a1/language-en

[135] WLTP 2017 *Worldwild Harmonised Light Vehicle Test Procedure*

[136] Colclasure A M and Kee R J 2010 Thermodynamically consistent modeling of elementary electrochemistry in lithium-ion batteries *Electrochim. Acta* **55** 8960–73

[137] IEA 2020 *Global EV Outlook 2020* (Paris: IEA) https://iea.org/reports/global-ev-outlook-2020

[138] Bloomberg New Energy Finance 2018 *Electric Buses in Cities Driving Towards Cleaner Air and Lower CO_2* (https://about.bnef.com/blog/electric-buses-cities-driving-towards-cleaner-air-lower-co2/)

[139] https://iea.org/data-and-statistics/charts/new-electric-bus-registrations-in-china-2015–2019

[140] https://wri.org/insights/how-did-shenzhen-china-build-worlds-largest-electric-bus-fleet

[141] https://sustainable-bus.com/fuel-cell-bus/fuel-cell-bus-hydrogen/

[142] IRENA 2019 *Innovation Outlook: Smart Charging for Electric Vehicles* (Abu Dhabi: International Renewable Energy Agency) https://irena.org/publications/2019/May/Innovation-Outlook-Smart-Charging

[143] Tran-Quoc T and Le Pivert X 2016 Stochastic approach to assess impacts of electric vehicles on the distribution network *IEEE PES 3rd Innovative Smart Grid Technologies EUROPE (ISGT Europe)(October 2012) (Berlin, Germany)*
Tuan Tran Q and Van Linh Nguyen L V Integration of electric vehicles into industrial network: impact assessment and solutions *IEEE/PES, General Meeting(17–21 July) (Boston, MA)*

[144] Hautier G 2019 Finding the needle in the haystack: materials discovery and design through computational ab initio high-throughput screening *Comput. Mater. Sci.* **163** 108–16

[145] Curtarolo S, Hart G L W, Nardelli M B, Mingo N, Sanvito S and Levy O 2013 *The high-throughput highway to computational materials design Nat. Mater.* **12** 191–201

[146] Saal J E, Kirklin S, Aykol M, Meredig B and Wolverton C 2013 *Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD)* *JOM* **65** 1501–9

[147] Cheng L *et al* 2015 Accelerating electrolyte discovery for energy storage with high-throughput screening *J. Phys. Chem. Lett.* **6** 283–91

[148] Aykol M *et al* 2016 High-throughput computational design of cathode coatings for Li-ion batteries *Nat. Commun.* **7** 13779

[149] Hautier G *et al* 2011 Phosphates as lithium-ion battery cathodes: an evaluation based on high-throughput ab initio calculations *Chem. Mater.* **23** 3495–508

[150] Wallace S K, Bochkarev A S, van Roekeghem A, Carrasco J, Shapeev A and Mingo N 2021 Free energy of $(CoxMn_{1-x})_3O_4$ mixed phases from machine-learning-enhanced ab initio calculations *Phys. Rev. Mater.* **5** 035402

[151] Wallace S K, van Roekeghem A, Bochkarev A S, Carrasco J, Shapeev A and Mingo N 2021 Modeling the high-temperature phase coexistence region of mixed transition metal oxides from ab initio calculations *Phys. Rev. Res.* **3** 013139

[152] Gubaev K, Podryabinkin E V, Hart G L W and Shapeev A V 2019 Accelerating high-throughput searches for new alloys with active learning of interatomic potentials *Comput. Mater. Sci.* **156** 148–56

[153] Novikov I S, Gubaev K, Podryabinkin E V and Shapeev A V 2021 The MLIP package: moment tensor potentials with MPI and active learning *Mach. Learn.: Sci. Technol.* **2** 025002
[154] Chandesris M, Caliste D, Jamet D and Pochet P 2019 Thermodynamics and related kinetics of staging in intercalation compounds *J. Phys. Chem.* C **123** 23711–20
[155] Onsager L 1931 Reciprocal relations in irreversible processes I *Phys. Rev.* **37** 405–26
[156] Feng X, Ouyang M, Liu X, Lu L, Xia Y and He X 2018 Thermal runaway mechanism of lithium ion battery for electric vehicles: a review *Energy Storage Mater.* **10** 246–67
[157] Lopez-Haro M, Guetaz L, Printemps T, Morin A, Escribano S, Jouneau P H, Bayle-Guillemaud P, Chandezon F and Gebel G 2014 Three-dimensional analysis of Nafion layers in fuel cell electrodes *Nat. Commun.* **5** 5229
[158] Dupre N, Moreau P, De Vito E, Quazuguel L, Boniface M, Bordes A, Rudisch C, Bayle-Guillemaud P and Guyomard D 2016 Multiprobe study of the solid electrolyte interphase on silicon-based electrodes in full-cell configuration *Chem. Mater.* **28** 2557–72
[159] Leriche J B *et al* 2010 An electrochemical cell for operando study of lithium batteries using synchrotron radiation *J. Electrochem. Soc.* **157** A606–10
[160] Tonin G, Vaugham G B M, Bouchet R, Allouin F, Di Michiel M and Barchasz C Operando investigation of the lithium/sulfur battery system by coupled x-ray absorption tomography and x-ray diffraction computed tomography *J. Power Sources* **468** 228287
[161] Bianchini M, Leriche J B, Laborier J L, Gendrin L, Suard E, Croguennec L and Masquelier C 2013 A new null matrix electrochemical cell for rietveld refinements of in situ or operando neutron powder diffraction data *J. Electrochem. Soc.* **160** A2176–83
[162] Tardif S, Dufour N, Colin J F, Gebel G, Burghammer M, Johannes A, Lyonnard S and Chandesris M 2021 Combining operando x-ray experiments and modelling to understand the heterogeneous lithiation of graphite electrodes *J. Mater. Chem.* A **9** 4281–90
[163] Tripathi A M, Wei-Nien Su W N and Hwang B J 2018 In situ analytical techniques for battery interface analysis *Chem. Rev. Soc.* **47** 736–851
[164] Benayad A, Morales-Ugarte J E, Santini C C, Bouchet R and Operando X P S 2021 A novel approach for probing the lithium/electrolyte interphase dynamic evolution *J. Phys. Chem.* A **125** 1069–81
[165] Satija R, Jacobson D L, Arif M and Werner S A 2004 In situ neutron imaging technique for evaluation of water management systems in operating PEM fuel cells *J. Power Sources* **129** 238–45
[166] Xu F, Diat O, Gebel G and Morin A 2007 Determination of transverse water concentration profile through MEA in a fuel cell using neutron scattering *J. Electrochem. Soc.* **154** B1389–98
[167] Martinez N, Porcar L, Escribano S, Micoud F, Rosini S, Tengattini A, Atkins D, Gebel G, Lyonnard S and Morin A 2019 Combined operando high resolution SANS and neutron imaging reveals in situ local water distribution in an operating fuel cell *ACS Appl. Energy Mater.* **2** 8425–33
[168] IEA 2019 *The Future of Hydrogen* (Paris: IEA) https://iea.org/reports/the-future-of-hydrogen
[169] The Royal Society 2019 *Sustainable Synthetic Carbon Based Fuels for Transport* (https://royalsociety.org/topics-policy/projects/low-carbon-energy-programme/sustainable-synthetic-carbon-based-fuels-for-transport/)

[170] Geelen L H W T, Kamps P T W J and Olsthoorn T N 2017 From overexploitation to sustainable use, an overview of 160 years of water extraction in the Amsterdam dunes, the Netherlands *J. Coast Conserv.* **21** 657–68
[171] Paull J 2009 A century of synthetic fertilizer: 1909–2009 *Element.—J. Bio-Dynam. Tasmania* **94** 16–20
[172] Binetti R, Costamagna F M and Marcello I 2008 Exponential growth of new chemicals and evolution of information relevant to risk control *Ann. Ist Super Sanità* **44** 13–5
[173] Daley J 2017 Science is falling woefully behind in testing new chemicals *Smithsonian Magazine*
[174] United Nations Environment Assembly of the United Nations Environment Programme (UNEP) 2019 *Global Chemicals Outlook II: Summary for Policymakers* (Geneva, Switzerland: UNEP/EA.4/21, UNEP)
[175] Card M L, Gomez-Alvarez V, Lee W-H, Lynch D G, Orentas N J, Lee M T, Wong E M and Boethling R S 2017 History of EPI Suite™ and future perspectives on chemical property estimation in US Toxic Substances Control Act new chemical risk assessments *Environ. Sci.: Processes Impacts* **19** 203–12
[176] Matthies M, Solomon K, Vighi M, Gilman A and Tarazona J V 2016 The origin and evolution of assessment criteria for persistent, bioaccumulative and toxic (PBT) chemicals and persistent organic pollutants (POPs) *Environ. Sci.: Processes Impacts* **18** 1114–28
[177] Allen J G These toxic chemicals are everywhere—even in your body. And they won't ever go away, 'Opinions' article published in the *Washington Post* 02 January 2018 (https://washingtonpost.com/opinions/these-toxic-chemicals-are-everywhere-and-they-wont-ever-go-away/2018/01/02/82e7e48a-e4ee-11e7-a65d-1ac0fd7f097e_story.html) (Accessed 15 March 2021)
[178] Gouin T, Mackay D, Jones K C, Harner T and Meijer S N 2004 Evidence for the 'grasshopper' effect and fractionation during long-range atmospheric transport of organic contaminants *Environ. Pollut.* **128** 139–48
[179] Fiedler H, Kallenborn R, de Boer J and Sydnes L K 2019 The Stockholm convention: a tool for the global regulation of persistent organic pollutants *Chem. Int.* **41** 4–11
[180] Sontake A R and Wagh S M 2014 The phase-out of perfluorooctane sulfonate (PFOS) and the global future of aqueous film forming foam (AFFF), innovations in fire fighting foam *Chem. Eng. Sci.* **2** 11–4
[181] Holmquist P H, Schellenberger S, van der Veen I, Peters G M, Leonards P E G and Cousins I T 2016 Properties, performance and associated hazards of state-of-the-artdurable water repellent (DWR) chemistry for textile finishing *Environ. Intern.* **91** 251–24
[182] Brandsma S H, Koekkoek J C, van Velzen M J M and de Boer J 2019 The PFOA alternative substitute GenX detected in grass and leaves near a fluoropolymer manufactory plant in the Netherlands *Chemosphere* **220** 493–500
[183] Cousins I T *et al* 2019 The concept of essential use for determining when uses of PFASs can be phased out *Environ. Sci.: Processes Impacts* **21** 1803–15
[184] de Boer J and Stapleton H M 2019 Toward fire safety without chemical risk *Science* **364** 231–2
[185] Autrup H *et al* 2020 Human exposure to synthetic endocrine disrupting chemicals (S-EDCs) is generally negligible as compared to natural compounds with higher or comparable endocrine activity. How to evaluate the risk of the S-EDCs? *Arch.Toxicol.* **94** 2549–57
[186] Zoeller R T *et al* 2014 A path forward in the debate over health impacts of endocrine disrupting chemicals *Environ. Health* **13** 118

[187] Daughton C G and Ternes T A 1999 Pharmaceuticals and personal care products in the environment: agents of subtle change? *Environ. Health Persp.* **107** 906–10
[188] Legler J and van Straalen N 2018 Decision-making in a storm of discontent *Science* **360** 958–60
[189] Blaquière T, Smagghe G, van Gestel C A M and Mommaerts V 2012 Neonicotinoids in bees: a review on concentrations, side-effects and risk assessment *Ecotoxicol* **21** 973–92
[190] Walters R 2010 Toxic atmospheres air pollution, trade and the politics of regulation *Crit. Crim.* **18** 307–23
[191] Skoczynska E 2021 Development and appliucation of comprehensive chemical analytical; methods for the analysis of polyaromatic compounds *PhD Thesis* Vrije Universiteit, Amsterdam, The Netherlands p 190
[192] Hassaan M A and El Nemr A 2017 Health and environmental impacts of dyes: mini review *Am. J. Environ. Sci. Eng.* **1** 64–7
[193] Kdasi A, Idris A, Saed K and Guan C T 2004 Treatment of textile wastewater by advanced oxidation processes: a review *Global Nest. Int.* J **6** 222–30
[194] Mohan V, Rao C and Karthikeyan J 2002 Adsorptive removal of direct azo dye from aqueous phase onto coal based sorbents: a kinetic and mechanistic study *J. Hazard. Mater.* **90** 189–204
[195] Wallington T J, Srinivasan J, Nielsen O J and Highwood E J 2009 Greenhouse gases and global warming *Environmental and Ecological Chemistry* ed A Sablic (Oxford: Eolss Publ. Co. Ltd) vol 1
[196] Plummer L N and Busenberg E 2000 Chlorofluorocarbons *Environmental Tracers in Subsurface Hydrology* ed P G Cook and A L Herczeg (Boston, MA: Springer)
[197] Birmpili T 2018 Montreal Protocol at 30: the governance structure, the evolution, and the Kigali Amendment *C. R. Geosci.* **350** 425–31
[198] Molina M J and Rowland F S 1974 Stratospheric sink for chlorofluoromethanes: chlorine atom-catalysed destruction of ozone *Nature* **249** 810–2
[199] Crutzen P J 1970 The influence of nitrogen oxides on the atmospheric ozone content *R. Meteorol. Suc. Quart. J.* **96** 320–5
[200] Rowland F S 1996 Stratospheric ozone Depletion by chlorofluorocarbons (Nobel lecture) *Angew. Chem. Int. Ed. Engl.* **35** 1786–98
[201] McKenzie R, Bernhard G and Liley B *et al* 2019 Success of Montreal Protocol demonstrated by comparing high-quality UV measurements with 'World Avoided' calculations from two chemistry-climate models *Sci. Rep.* **9** 12332
[202] Montzka S A, Velders G J M, Krummel P B, Mühle J, Orkin V L, Park S, Shah N, H and Walter-Terrinoni 2018 Hydrofluorocarbons *Scientific Assessment of Ozone Depletion, Ch. 2: Global Ozone Research and Monitoring Project-Report No. 58* (Geneva: Chemical Sciences Laboratory)
[203] Flerlage H, Velders G J M and de Boer. J 1992 A review of bottom-up and top-down emission estimates of hydrofluorocarbon(HFCs) in different parts of the worlds *Chemosphere* **283** 131208
[204] Fujika M and Tajima S The pollution of Minamata Bay by mercury *Water Sci. Technol.* **25** 133–40
[205] de Boer M A and Lammertsma K 2013 Scarcity of reare earth elements *ChemSusChem* **6** 2045–55 2013

[206] Boekhout van Solinge T 2008 Crime, conflicts and ecology in Africa *Global Harms: Ecological Crime and Speciesism* ed T Sollund (Nova Science Publ.) pp 13–35
[207] Aoun M, El Samrani A G, Lartiges B S, Kazpard V and Saad Z 2010 Releases of phosphate fertilizer industry in the surrounding environment: Investigation on heavy metals and polonium-210 in soil *J. Environ. Sci.* **22** 1387–97
[208] Zagà V, Lygidakis C, Chaouachi K and Gattavecchia E 2011 Polonium and lung cancer *J. Oncol.* **2011** 860103
[209] Day R H 1980 The occurrence and characteristics of plastic pollution in Alaska's marine birds *M.S. Thesis* University of Alaska. Fairbanks, AK, USA p 111
[210] Moore C J, Moore S L, KLeecaster M and Weisberg S B 2001 A comparison of plastic and Plankton in the North Pacific Central Gyre *Mar. Pollut. Bull.* **42** 1297–300
[211] Law K L 2017 Plastics in the environment *Ann. Rev. Mar. Sci.* **9** 205–29
[212] Landrigan P 2017 Air pollution and health *Lancet* **2** E4–5
[213] Lelieveld J, Evans J S, Fnais M, Giannadaki D and Pozzer A 2015 The contribution of outdoor air pollution sources to premature mortality on a global scale *Nature* **525** 367–71
[214] Kriebel D, Tickner J, Epstein P, Lemons J, Levins R, Loechler E L, Quinn M, Rudel R, Schettler T and Stoto M 2001 The precautionary principle in environmental science *Environ. Health Perspect.* **109** 871–6
[215] Hendriks H S, Meijer M, Muilwijk M, Van den Berg M and Westerink R H S 2014 A comparison of the *in vitro* cyto- and neurotoxicity of brominated and halogen-free flame retardants: prioritization in search for safe(r) alternatives *Arch. Toxicol.* **88** 857–69
[216] Hák T, Janoušková S and Moldan B 2016 Sustainable development goals: a need for relevant indicators *Ecol. Indic.* **60** 565–73
[217] Stafford-Smith M, Griggs D, Gaffney O, Ullah F, Reyers B, Kanie N, Stigson B, Shrivastava P, Leach M and O'Connell D 2017 Integration: the key to implementing the sustainable development goals *Sustain. Sci.* **12** 911–9
[218] EPA 2019 Chemical safety for sustainability national research program strategic research action plan 2019–2022 *Office of Research and Development Chemical Safety for Sustainability, No. CSS FY2019-FY2022 StRAP* (Washington DC: EPA)
[219] https://www.mondaq.com/china/chemicals/775388/china-publishes-draft-regulation-on-the-environmental-risk-assessment-and-control-of-chemical-substances-the-real-china-reach
[220] Robinson R M and Behnke R A 2001 *The U.S. National Space Weather Program: A retrospective* (Washington, DC: American Geophysical Union Geophysical Monograph Series) 125 10
[221] Hapgood M A 2011 Towards a scientific understanding of the risk from extreme space weather *Adv. Space Res.* **47** 2059–72
[222] Schrijver C J *et al* 2015 Understanding space weather to shield society: a global road map for 2015–2025 commissioned by COSPAR and ILWS *Adv. Space Res.* **55** 2745–807
[223] Green L and Baker D 2015 Coronal mass ejections: a driver of severe space weather *Weather* **70** 31–5
[224] Richardson I G, Cliver E W and Cane H V 2001 Sources of geomagnetic storms for solar minimum and maximum conditions during 1972–2000 *Geophys. Res. Lett.* **28** 2569–72
[225] National Academies Press 2008 *Severe Space Weather Events–Understanding Societal and Economic Impacts* (National Academies Press)
[226] Kreutzer R 2009 Space weather *Global Risk Dialogue* (Berlin: Springer) pp 12–7

[227] Eastwood J P, Biffis E, Hapgood M A, Green L, Bisi M M, Bentley R D, Wicks R, McKinnell L A, Gibbs M and Burnett C 2017 The economic impact of space weather: where do we stand *Risk Anal.* **37** 206–18
[228] Cannon P S 2013 Extreme space weather—a report published by the UK Royal Academy of Engineering *Space Weather* **11** 138–9
[229] Pete R 2012 On the probability of occurrence of extreme space weather events *Space Weather* **10** 02012
[230] Schrijver C J, Dobbins R, Murtagh W and Petrinec S M 2014 Assessing the impact of space weather on the electric power grid based on insurance claims for industrial electrical equipment *Space Weather* **12** 487–98
[231] Howard J W Jr and Hardage D M 1999 Spacecraft environments interactive: space radiation and its effects on electronic system *Technical Report, NASAs* https://ntrs.nasa.gov/citations/19990116210
[232] https://euweather.eu/stations_stats.php?en
[233] Odstrcil D 2003 Modeling 3-D solar wind structure *Adv. Space Res.* **32** 497–506
[234] Pomoell J and Poedts S 2018 EUHFORIA: European heliospheric forecasting information asset *J. Space Weather Space Clim.* **8** A35
[235] Poedts S *et al* 2020 The virtual space weather modelling centre *J. Space Weather Space Clim.* **10** 14
[236] Vainio R and Laitinen T 2007 Monte Carlo simulations of coronal diffusive shock acceleration in self-generated turbulence *Astrophys. J.* **658** 622–30
[237] Afanasiev A, Battarbee M and Vainio R 2015 Self-consistent Monte Carlo simulations of proton acceleration in coronal shocks: effect of anisotropic pitch-angle scattering of particles *Astron. Astrophys.* **584** A81
[238] Pomoell J, Aran A, Jacobs C, Rodrguez-Gasén R, Poedts S and B S 2015 Modelling large solar proton events with the shock-and-particle model. Extraction of the characteristics of the MHD shock front at the cobpoint *J. Space Weather Space Clim.* **5** A12
[239] Wijsen N, Aran A, Sanahuja B, Pomoell J and Poedts S 2020 The effect of drifts on the decay phase of SEP events *Astron. Astrophys.* **634** A82
[240] Wijsen N, Samara E, Aran À, Lario D, Pomoell J and Poedts S 2021 A self-consistent simulation of proton acceleration and transport near a high-speed solar wind stream *Astrophys. J. Lett.* **908** L26
[241] Lakka A, Pulkkinen T I, Dimmock A P, Kilpua E, Ala-Lahti M, Honkonen I, Palmroth M and Raukunen O 2019 GUMICS-4 analysis of interplanetary coronal mass ejection impact on Earth during low and typical Mach number solar winds *Ann. Geophys.* **37** 561–79
[242] Eggington J W B, Eastwood J P, Mejnertsen L, Desai R T and Chittenden J P 2020 Dipole tilt effect on magnetopause reconnection and the steady-state magnetosphere-ionosphere system: global MHD simulations *J. Geophys. Res. (Space Phys.)* **125** e27510
[243] Pierrard V and Voiculescu M 2011 The 3D model of the plasmasphere coupled to the ionosphere *Geophys. Res. Lett.* **38** L12104
[244] Bourdarie S A and Maget V F 2012 Electron radiation belt data assimilation with an ensemble Kalman filter relying on the Salammbô code *Ann. Geophys.* **30** 929–43
[245] cf http://ccmc.gsfc.nasa.gov/
[246] https://clasp.engin.umich.edu/research/theory-computational-methods/center-for-space-environment-modeling/

[247] Raeder J, Larson D, Li W, Kepko E L and Fuller-Rowell T 2008 OpenGGCM simulations for the THEMIS mission *Space Sci. Rev.* **141** 535–55

[248] Opgenoorth H J *et al* 2019 Assessment and recommendations for a consolidated European approach to space weather—as part of a global space weather effort *J. Space Weather Space Clim.* **9** A37

[249] Scolini C, Rodriguez L, Mierla M, Pomoell J and Poedts S 2019 Observation-based modelling of magnetised coronal mass ejections with EUHFORIA *Astron. Astrophys.* **626** A122

[250] Scolini C *et al* 2020 CME-CME interactions as sources of CME geoeffectiveness: the formation of the complex ejecta and intense geomagnetic storm in 2017 Early September *Astrophys. J. Suppl.* **247** 21

[251] Kilpua E K J, Olspert N, Grigorievskiy A, Käpylä M J, Tanskanen E I, Miyahara H, Kataoka R, Pelt J and Liu Y D 2015 Statistical study of strong and extreme geomagnetic disturbances and solar cycle characteristics *Astrophys. J.* **806** 272

[252] Horwitz J L, Gallagher D L and Peterson W K 1998 *Geospace Mass and Energy Flow: Results from the International Solar-Terrestrial Physics Program* (Washington, DC: American Geophysical Union Geophysical Monograph Series) p 104

[253] Lilensten J *et al* 2021 Quo vadis, European Space Weather community? *J. Space Weather Space Clim.* **11** 26

[254] Antiochos S K, DeVore C R and Klimchuk J A 1999 A model for solar coronal mass ejections *Astrophys. J.* **510** 485–93

[255] Lodders K 2003 Solar system abundances and condensation temperatures of the elements *Astrophys. J.* **591** 1220–47

[256] Reames D V 1999 Particle acceleration at the Sun and in the heliosphere *Space Sci. Rev.* **90** 413–91

[257] Daly E J 2001 *ESA Space Weather Activities* ed I A Daglis (Hazards: Space Storms and Space Weather) p 459

[258] https://cordis.europa.eu/