
CHAPTER 1

Physics Overview of Solar Energy

DIEGO CASTANO

1.1 INTRODUCTION

Undoubtedly the most important factor in the study of solar energy is the sun, the local star and the gravitational stake to which the earth is tethered. All forms of energy on earth, except for nuclear, can ultimately be traced to the sun. There are on the order of a hundred billion stars in the Milky Way galaxy and about a hundred billion galaxies in the known universe. Stars began forming several hundred thousand years after the Big Bang, which, based on current theories of cosmology, happened 13.7 billion years ago. The sun itself formed about 5 billion years ago and should shine for another 5 billion. The earth consequently finds itself in a somewhat special place in space and time. The low entropy state of the solar system ensures vital change for billions of years to come. In terms of thermodynamics, the sun represents an effective high temperature reservoir of temperature 6000 kelvins that bathes the earth (the low temperature reservoir at an average 287 kelvins) with a radiant intensity of approximately 1400 W/m^2 above the atmosphere. Attenuation due to absorption and scattering in the atmosphere reduces this value. The exact value of solar irradiance at the earth's surface depends on the sunlight's path through the atmosphere and is generally no greater than 1000 W/m^2 . The solar irradiance translates into a naive, maximal power output of approximately 10^{17} watts (compare this to the world's average energy consumption rate in 2010, of just over 10^{13} watts [1]). It would appear that the sunlight reaching just earth (the earth subtends a mere $4.6 \times 10^{-8}\%$ of the whole solid angle at the sun; see Fig. 1.1) can supply earth's needs 10,000 times over. In fact, Freeman Dyson considered the possibility that advanced civilizations would surround their suns with enough orbiting artificial satellites (this system is referred to as a Dyson sphere) to harness most of the star's solar energy. In the case of the sun, an optimal Dyson sphere would generate on the order of 10^{27} watts, a significant fraction of the luminosity of the sun ($L_{\odot} = 4.8 \times 10^{27} \text{ W}$).

2 PHYSICS OVERVIEW OF SOLAR ENERGY

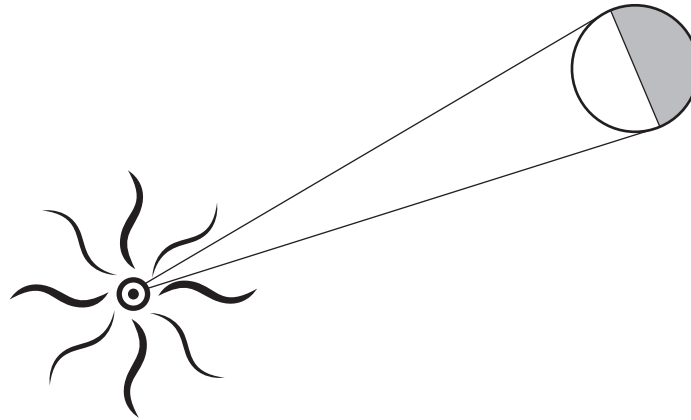


FIGURE 1.1 The earth subtends six nanosteradians at the sun.

In fact, about 30% of the sunlight reaching the earth is reflected back into space above the atmosphere. The albedo, or fraction of reflected sunlight, depends strongly on such factors as snow and cloud cover. The 70% of the incident sunlight not reflected is effectively transmitted through the atmosphere, which is mostly transparent to visible light, to the surface. The thermal, or infrared, radiation emitted by the surface of the earth is then absorbed by the atmosphere and reradiated. This atmospheric greenhouse effect is what leads to a global mean surface temperature of 287 kelvins as compared to a freezing 254 kelvins without an atmosphere.

1.2 THE SUN

The standard theories of particle physics and cosmology describe the Big Bang as the moment of creation of space, time, matter, and energy. At first, there was just one superforce to mediate the interactions between particles. The superforce eventually separated out into the four known fundamental forces of today, that is, gravity, electromagnetism, weak and strong nuclear. All four of these forces are important in the development and function of the sun.

Early on (about 20 minutes after the Big Bang) the universe is a cauldron containing, among other particles, electrons and various light nuclei. Recombination, which occurs around 300,000 years after the Big Bang, is the term used to describe the moment in the history of the universe when electromagnetically neutral bound states of matter were first possible in large numbers. In this era, the universe is a rarified gas consisting mostly of hydrogen (75%) and helium (25%). Gravity starts to make its presence known and leads to clumping due to inhomogeneities in the matter distribution. As a hydrogen/helium gas clump shrinks, the gravitational energy is turned into kinetic energy, resulting in temperatures at the core high enough to fuse the hydrogen and helium. At a temperature of ten million kelvins, protons can tunnel

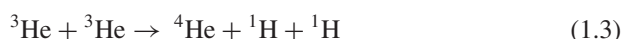
through the Coulomb repulsion barrier, resulting in the first nuclear reaction of the so-called proton–proton cycle, which is prevalent in stars like the sun,



Note that the fusing of the protons involves the strong nuclear force, and the appearance of the neutrino indicates that the weak nuclear force was also involved in this process. This reaction is followed by the reaction



The last reaction in the cycle can be any one of four with the most common in the sun being



The net energy released along this sequence is 27 MeV. This is only one sequence of many possible ones, all of which lead to ${}^4\text{He}$. In stars like the sun this eventually leads to an inert helium core with the hydrogen fusion occurring in the shell surrounding the core. This causes the star to grow and become a red giant. The helium core continues to collapse gravitationally until the temperature increases to the point at which helium–helium fusion can occur. The fusion of heavier and heavier nuclei continues in stars that are massive enough until ${}^{56}\text{Fe}$ is reached. This isotope of iron has the largest binding energy per nucleon, and fusion reactions that produce heavier nuclei are consequently endothermic. Although heavy elements can still be produced in stars, most heavy elements are created in supernovas.

One way astronomers classify stars is by their luminosities and spectral types or surface temperatures. The resulting scatter plot is called a Hertzsprung–Russell diagram (Fig. 1.2). Most stars, including the sun, fall in the region referred to as the main sequence.

1.3 LIGHT

The history of the developments in the theory of light is a long one. The Greeks, from as far back as Pythagoras, believed that light emanated from visible bodies. Some even philosophized that it traveled at a finite speed. However, most people before the 17th century believed that light was instantaneous. Galileo is credited with being the first well-known scientist to attempt to measure the speed of light. His experiment involved him and an assistant on distantly separated hills with lanterns and some sort of time measuring device, perhaps a water clock. Due to the inherent lack of precision in the design of the experiment, his result was ambiguous. Of the speed of light, he is alleged to have said, “If not instantaneous, it is extraordinarily rapid.” He also concluded that the speed was at least ten times faster than sound. About a decade later, Ole Rømer used essentially the same idea as Galileo—that of measuring the

4 PHYSICS OVERVIEW OF SOLAR ENERGY

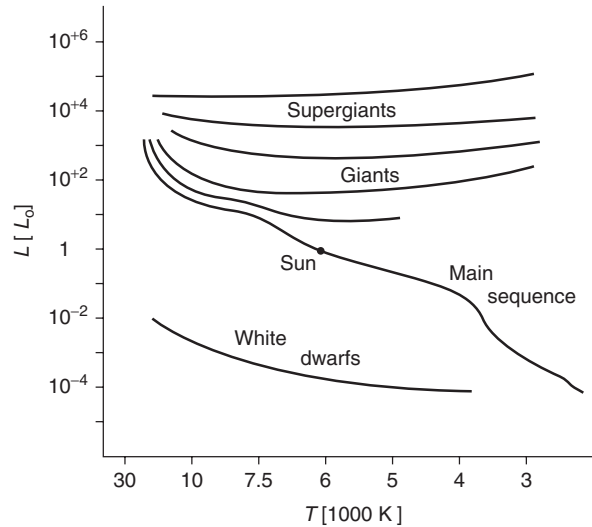


FIGURE 1.2 Sketch of Hertzsprung–Russell diagram.

time it takes a light signal to cover a spatial path—but Rømer did it with a longer path, the diameter of the earth’s orbit around the sun. His results were considerably better ($c = 2 \times 10^8$ m/s). It took about another 200 years to perfect an earth bound experiment, but Leon Foucault performed an experiment involving rotating mirrors (based on one designed by Hippolyte Fizeau using rotating toothed wheels) that was able to measure the speed of light accurately to good precision and that agrees with modern measurement to four significant figures ($c = 2.998 \times 10^8$ m/s).

The nature of light, whether it be a particle or wave phenomenon, was a topic of debate in the 17th century. Isaac Newton was on the particle side and referred to the constituent particles as corpuscles. On the wave side was Christian Huygens. Newton’s reputation helped the particle side, but the phenomena of interference and diffraction weighed heavily on the side of the wave theory. Today, at least in classical terms, light is recognized as a wave phenomenon, but it is interesting to note that quantum mechanics has forced reconsideration and Newton’s corpuscles can be thought resurrected as photons.

The unification of electricity and magnetism made by James Clerk Maxwell in the mid-19th century put the wave nature of light on firm theoretical footing. Maxwell’s correction to Ampere’s law led him to wave equations for electric and magnetic fields with a wave speed relation that involved electric permittivity and magnetic permeability constants

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \quad (1.4)$$

This calculable wave speed coincided with the experimentally measured speed of light. The conclusion was inescapable: light is a manifestation of electromagnetic

waves. In 1800 Frederick William Herschel had ascertained the existence of an invisible form of light by noting that the shadow region beyond the red end of a prism-induced spectrum of sunlight registered a higher temperature than the red lighted region. This invisible light is recognized today as the infrared.

Although it is far afield from the main thrust of this overview, it is perhaps interesting to note that questions immediately arose concerning the nature of the supporting medium for these waves. Whatever the medium was, it pervaded all of space and was referred to as the ether. The possibility of anisotropies in the speed of light due to the ether would occupy the minds of theorists and the efforts of experimentalists, in particular, Albert Michelson and Edward Morley, until the beginning of the 20th century. The null result of the late 19th century Michelson–Morley experiment to detect the so-called ether wind was, in retrospect, consistent with Albert Einstein’s theory of relativity.

Along with the theory of relativity, the early 20th century saw the development of quantum mechanics. The understanding of various light phenomena was the catalyst for its inception. The story begins with the concept of a blackbody, an idealized body capable of absorbing all incident electromagnetic radiation. This is to be compared with a real body, which reflects and/or transmits some fraction of the incident radiation. A perfect absorber is also a perfect emitter by Kirchhoff’s law of radiation, so a blackbody is also a perfect emitter. A kiln with a small opening is a good realization of a blackbody emitter. The radiation emitted was measured (using a bolometer), and two important facts were discovered. The intensity radiated (or total radiant emittance) at all wavelengths depends only on the absolute surface temperature of the blackbody according to the Stefan–Boltzmann law

$$I_b = \sigma T^4 \quad (1.5)$$

where $\sigma = 5.67 \times 10^{-8} \text{ W/(m}^2 \cdot \text{K}^4)$ is the Stefan–Boltzmann constant. Also discovered was the fact that there was a peak to the intensity distribution, known as Wien’s law, at

$$\lambda = \frac{0.003 \text{ m} \cdot \text{K}}{T} \quad (1.6)$$

Max Planck set out to theoretically derive the blackbody intensity distribution using the theory of electromagnetism and the laws of thermodynamics. He met with failure until he hypothesized the radiation quantum of energy

$$E = hf \quad (1.7)$$

where $h = 6.63 \times 10^{-34} \text{ J} \cdot \text{s}$ and f is the frequency of the radiation. With this assumption, he was able to derive Planck’s law of blackbody radiation (see Fig. 1.3):

$$\frac{dI_b(\lambda, T)}{d\lambda} = \frac{2\pi hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda k_B T} - 1} \quad (1.8)$$

6 PHYSICS OVERVIEW OF SOLAR ENERGY

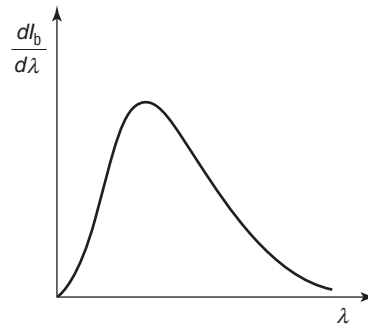


FIGURE 1.3 Characteristic blackbody curve.

By fitting the sun's radiation curve to this formula, its surface temperature can be deduced to be 6000 kelvins. The earth's own effective radiant temperature is around 240 kelvins.

Planck's idea was applied by Einstein to explain the photoelectric effect. In the late 19th century experiments conducted by Heinrich Hertz and Philipp Lenard showed that the energy of electrons ejected from a metal upon electromagnetic irradiation was independent of the radiation's intensity but depended on its frequency. By assuming that the electromagnetic radiation transferred only a quantum of energy to the electrons in the metal, Einstein was able to explain the effect. The predicted maximum energy of a free photoelectron was

$$KE_{\max} = hf - \Phi \quad (1.9)$$

where Φ is the work function, the minimum energy required to strip the electron from the metal. This is the basis of the photoemissive cell or phototube (see Fig. 1.4).

1.4 THERMODYNAMICS

It is often stated that thermodynamics was developed in response to the desire to make a better steam engine. As such it can be thought of as the study of thermal energy conversion. So what is energy? The term is pervasive in the modern world,

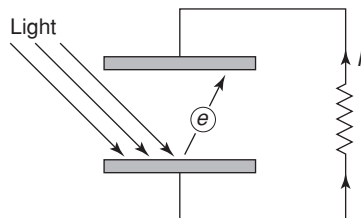


FIGURE 1.4 A current produced by the photoelectric effect.

and the concept is often reified, especially in scientifically informal settings. Energy is, in fact, an abstract concept, and its conservation (constancy in time) is a powerful organizational principle in physics. It comes in an apparent myriad of forms, such as solar, chemical, electrical, and even dark, but they can all be placed into two broad categories, kinetic and potential. Perhaps the most intuitive, if not the most perceptively apparent, is the kinetic type: If there is motion in a system, then there is energy. Conversely, and naively to be sure, energy is what makes things go.

There is (thermodynamic) energy associated with matter due to its atomic nature and resulting from the randomized (thermal) motion and from the interactions of the constituent particles. One of the most important results from the kinetic theory of gases is that temperature is a measure of the average kinetic energy of the constituents of matter. The finite size of bodies implies that the kinetic energies of their constituents must be changing due to collisions and oscillations. It is known from atomic theory that matter contains charged particles, such as electrons and protons. Maxwell's electromagnetic theory predicts that accelerated charged particles will radiate. Therefore all bodies at nonzero absolute temperature should radiate, and Eq. 1.8 predicts the spectrum of an ideal one.

In thermodynamics, the universe is divided into two parts, the system under consideration and its surroundings. All exchanges between system and surroundings are done across a boundary, the real or effective surface separating the two. The total energy of a thermodynamic system is referred to as its internal energy (U) and includes kinetic energies as well as the potential energies of particle interactions. In thermodynamics, heat (Q) is the process variable that represents the microscopically unobservable transfer of energy, whereas work (W) represents the macroscopically observable transfer of energy. Heat transfer is typically driven by diathermal, or unrestricted, contact between two bodies of differing temperatures. A boundary that does not permit the flow of heat is called adiabatic. The first law of thermodynamics is the statement of energy conservation. Any changes in the internal energy of a system must be the result of energy transfer across the boundary,

$$dU = d'Q - d'W \quad (1.10)$$

where d' implies an inexact differential, and the work is considered done by the system.

Colloquially, the task of building an engine, for example, a steam engine, amounts to devising a way to turn internal into external energy, or how to use a source of heat to make something go. The first steam engine (the aeolipile; see Fig. 1.5) is credited to Hero of Alexandria, who lived in the 1st century A.D. A tropical hurricane is a natural example of a steam engine, effectively using the thermal energy in the surface water of the ocean to power wind. James Watt perfected the modern steam engine in the late 18th century. In all engines, some energy (in fact, generally quite a bit) is always wasted and in a form that is useless to the engine. The design details of the engine will affect the exact amount of wasted energy but there is always a nonzero amount, as surely as heat always flows spontaneously from higher to lower temperature. The second law of thermodynamics is the statement that the preceding empirical

8 PHYSICS OVERVIEW OF SOLAR ENERGY

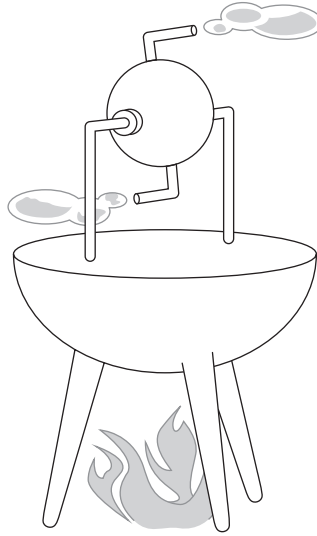


FIGURE 1.5 The aeolipile.

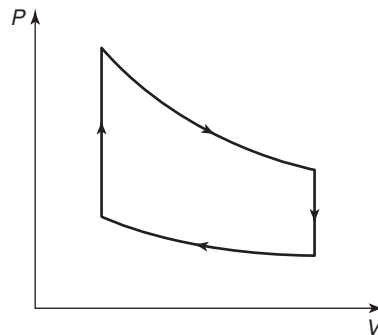
observation is indeed the case. The second law can be put into a mathematical form by noting that there exists a function of the extensive parameters of a system in a state of equilibrium that is maximized in any spontaneous process. The function is called the entropy (S) and for quasistatic processes

$$dS = \frac{d'Q}{T} \quad (1.11)$$

Engines exploit a temperature difference to extract useful work. The ideal cyclic engine, given two working temperatures, is one for which the entropy change of the universe is zero (i.e., its operation is reversible). Such an engine is called a Carnot engine, named after Sadi Carnot who first conceptualized it. The Carnot cycle consists of two isothermal processes and two adiabatic ones. The Carnot engine's efficiency depends on the two reservoir temperatures,

$$\varepsilon_{\text{Carnot}} = 1 - \frac{T_L}{T_H} \quad (1.12)$$

where $T_{L(H)}$ is the lower (higher) heat reservoir temperature. Unfortunately, the Carnot engine is an idealization. No heat transfer process is ever reversible in practice. Moreover, the sequence of processes associated with the Carnot cycle would be difficult to realize in a practical way. A more practical engine than Carnot's, with a theoretical efficiency that nevertheless matches Carnot's, is the Stirling engine (with regenerator). It is an external combustion engine that uses a single phase working substance (a gas, such as air). The Stirling cycle consists of two isothermal and two isochoric processes (see Fig. 1.6). Although they generally have low power outputs

**FIGURE 1.6** Stirling cycle.

for their size, Stirling engines are relatively easy to make and can exploit even small temperature differences.

There are at least four obvious ways to use direct energy from the sun. The first involves the direct absorption of sunlight. Architecturally, living spaces can be warmed by designing them to take advantage of sunlight. Water can also be heated directly by the sun for different uses such as space heating. By using devices, such as Fresnel lenses or parabolic mirrors, the sun's rays can be concentrated to increase input. Fresnel lenses have large aperture and dioptric power. Parabolic mirrors have the property that incident parallel light is focused without aberration. Both of these can collect light over a relatively large area and intensify it significantly.

The second way to exploit the thermal energy from the sun is to use an engine, perhaps in conjunction with some focusing system such as the ones discussed above. The Stirling engine can easily be adapted to exploit an external heat source and has consequently gained popularity in the solar energy business. Thermoelectric generators based on the Seebeck, or thermoelectric, effect can also be used to convert thermal energy into electricity, although these are usually less efficient than Stirling engines.

The third way is the direct conversion of sunlight into electricity and is the subject of Section 1.5.

1.5 PHOTOVOLTAICS

As discussed above, photoemission can be used to generate a current, but there is another related photovoltaic effect that involves semiconductors rather than metals. The discovery of the photovoltaic effect is credited to Alexandre-Edmond Becquerel in 1839. He discovered that illuminating one of the electrodes in an electrolytic cell caused the current to increase. The effect was also seen in solids, like selenium, in the late 19th century. In 1954, Bell Laboratories produced the first photovoltaic cell using a crystalline silicon semiconductor. In the photovoltaic effect, like in the photoelectric

10 PHYSICS OVERVIEW OF SOLAR ENERGY

effect, light energy is absorbed by electrons, but, unlike in the photoelectric effect, the electrons are not ejected from the semiconductor.

Crystalline silicon, for example, has an extended, regular atomic pattern called a lattice. The basic unit of the lattice is a group of five silicon atoms. They form a geometrical array in which one is at the center of a tetrahedron and the other four at the centers of its faces. This structure is a consequence of the four outer valence electrons in silicon's third, or M, shell. When many identical atoms in close proximity are considered together, such as in a lattice, their electron's otherwise discrete energies (i.e., in isolation) are smeared into bands. In insulators, the highest band that is occupied by electrons at absolute zero is called the valence band. The next higher band, which is empty in this case, is called the conduction band. The energy gap between these bands in insulators is relatively large compared to characteristic thermal energies (0.025 eV at $T = 300$ K), so even above absolute zero an insulator has an effectively empty conduction band. In a conductor, the conduction band contains electrons even at absolute zero. In simplest terms, the electrons in the conduction band can be considered effectively free, that is, not bound to any one atom and therefore able to roam throughout the conductor. The electrons that are freed from covalent bonds leave a positively charged hole in the fixed lattice substratum. When a hole is filled by a neighboring electron, it in turn leaves behind a hole. The effective movement of the hole resembles that of a positively charged particle; the effect is similar to that of bubbles in a fluid. In a semiconductor, at absolute zero, the conduction band is empty like for an insulator. However, the energy gap is significantly smaller than in an insulator, and as the temperature rises thermal agitation is sufficient to lift electrons into the conduction band, making the material a conductor. Silicon has an energy gap of 1.1 eV at 300 K. In fact, every electron in the conduction band is accompanied by a hole in the valence band.

To change the balance of electrons in the conduction band and of holes in the valence band, a semiconducting material must be doped; that is, an impurity must be introduced into the otherwise homogeneous lattice. If phosphorus is introduced into the silicon lattice it will form the same four covalent bonds, but it will have an extra loosely bound electron. The energy level of these extra electrons lies in the energy gap just below the conduction band, so they readily become conduction electrons leaving behind a fixed positive ion. Impurities that have an extra valence electron, like phosphorus, are called donors. The semiconductor produced by donor doping is called n-type since the majority charge carriers are negative electrons. In an analogous fashion, silicon can be doped with an impurity that has one less valence electron, like boron. This type of impurity is called an acceptor. In this case the boron takes an electron from a neighboring silicon atom and creates a hole in the valence band. The energy level of these stolen (acceptor) electrons is just above the valence band. The holes are free to move as positive charge carriers, and the negative boron ions are fixed. The semiconductor produced by acceptor doping is called p-type since positive holes are the majority charge carriers.

Light incident on a semiconductor will generate electron-hole pairs if the photon energy is greater than the band gap. However, the pair will recombine unless the two carriers can be kept separated. Therefore to generate electrical power from incident

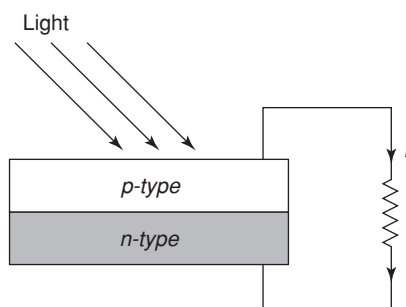


FIGURE 1.7 A solar cell in action.

light on these semiconductor materials requires a potential barrier. This barrier can be set up through the juxtaposition of p- and n-type materials. At the junction, both conduction and valance electrons move from the n-type side to the p-type side, thereby filling holes in the p-side and creating holes in the n-side. This process is finite and happens quickly. It creates a net fixed positive charge on the n-side of the junction and a net negative fixed charge on the p-side. This is known as the depletion zone and represents the potential barrier. It stops any further flow of electrons into the p-side and of holes into the n-side, but it does not prevent electrons from moving into the n-side and holes from moving into the p-side. Ideally a light-generated electron–hole pair will be separated by this mechanism and an emf generated.

If the two sides of the solar cell are connected through a load, a current will flow (see Fig. 1.7). Unfortunately, not all electron–hole pairs generated in this way will contribute to the current. The probability that a pair will contribute depends on many factors, one of which is the location within the cell where the pair forms. Those pairs created in the depletion zone will separate with certainty, then with high probability avoid combination and contribute to the current. Efficiency is typically a measure of output to input ratio. In the case of a solar cell, the input is solar energy and the output is electrical energy. The efficiency of a solar cell depends on many factors such as the one just described. Among other things, it also depends on the fraction of incident light absorbed and the nature (intensity and spectrum) of that light. The selenium cells of the late 19th century had 1% efficiency. The 1954 Bell Labs silicon cell had 4% efficiency. The maximum theoretical efficiency of a single p-n junction solar cell is given by the Shockley–Queisser limit of 31%. By using focusing devices and multijunction cells, that efficiency can be increased to 41%. A naive application of Eq. 1.12 would set an absolute upper limit of 95%. A more sophisticated calculation yields an upper limit of 86.6% [2].

1.6 PHOTOSYNTHESIS

This book is focused on the topic of photosynthesis. In this way, light energy generates an electron flow, which results in the production of energetic molecules that make

12 PHYSICS OVERVIEW OF SOLAR ENERGY

reduced organic compounds. The photosynthetic mechanism absorbs light energy by using pigments, especially chlorophyll (Chl) molecules. Photosynthetic organisms only need 1% of the solar spectrum to provide enough biomass and oxygen to support life on earth. Two photosystems work in tandem to carry out oxygenic photosynthesis. They are very efficient systems, using most of their pigments as antennas to harvest the light energy and then to transfer it to a very few, special Chl molecules. Photochemistry begins when these special Chl molecules donate electrons, which travel through the electron transport chain, reaching nicotinamide adenine dinucleotide phosphate (NADP^+) at photosystem I. NADPH is used for CO_2 fixation into organic compounds. The ultimate electron donor in oxygenic photosynthesis is water, which is oxidized to oxygen by photosystem II.

For more detailed accounts of the subjects discussed in this overview, see [3–6], and [7].

REFERENCES

1. British Petroleum, Statistical Review of World Energy 2011, <http://www.bp.com/sectionbodycopy.do?categoryId=7500&contentId=7068481> (accessed September 1, 2011).
2. M. A. Green, *Third Generation Photovoltaics: Advanced Solar Energy Conversion*, Springer, New York, 2006.
3. J. W. Rohlfs, *Modern Physics from α to Z_0* , 1st ed., Wiley, Hoboken, NJ, 1994.
4. M. L. Kutner, *Astronomy: A Physical Perspective*, 2nd ed., Cambridge University Press, Cambridge, U.K., 2003.
5. E. Hecht, *Optics*, 4th ed., Addison-Wesley, Reading, MA, 2002.
6. H. B. Callen, *Thermodynamics: An Introduction to the Physical Theories of Equilibrium Thermodynamics and Irreversible Thermodynamics*, Wiley, Hoboken, NJ, 1960.
7. C. Kittel, *Introduction to Solid State Physics*, 8th ed., Wiley, Hoboken, NJ, 2005.