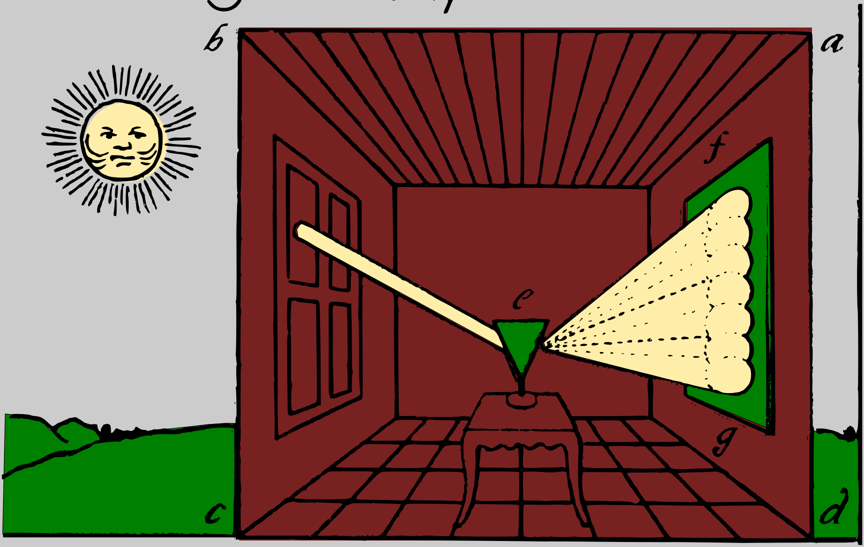
Book 4

Physics for Everyone

A. I. Kitaigorodsky



PHOTONS AND NUCLEI

Mir Publishers Moscow

Physics for Everyone Book 4

The book concludes the series Physics for Everyone by the world-renowned scientist, winner of the Nobel and Lenin prizes, academician Lev Landau and the distinguished physicist Alexander Kitaigorodsky. This book discusses in a simple and easy-to-understand manner the i phenomena of electromagnetic waves, thermal radiation, and current treatment of spectroscopic analysis. Provides an introduction to the field of nuclear physics and explains the most common types of lasers. Outlines principal aspects of special theory of relativity, and quantum mechanics.

ИЗДАТЕЛЬСТВО · МИР ·

Physics for Everyone

Book 4

A. I. Kitaigorodsky

PHOTONS AND NUCLEI

Translated from
the Russian
by George Yankovsky

Mir Publishers Moscow

Физика для
Книга
А. Китайгородский
Фотоны ядра
Издате. «Наука»

First published 1981

На английском язык

PREFACE

This is the fourth and concluding book in the *Physics for Everyone* series and it deals with the fundamentals of physics.

"Fundamentals" is of course a rather vague word but we will think of it as meaning the general laws on which the whole edifice of modern physics rests. There are not so many of them, and so we can make a list: the laws of motion of classical mechanics, the laws of thermodynamics, the laws that are embodied in the equations of Maxwell and that govern charges, currents and electromagnetic fields, and then the laws of quantum physics and the theory of relativity.

The laws of physics, like those of natural science at large, are of an empirical nature. They are arrived at by means of observation and experiment. Experiments establish a multitude of primary facts such as the building up of matter from atoms and molecules, the nuclear model of the atom, the wave-particle aspect of matter, and so on. Now, both the number of basic laws and also the number of fundamental facts and concepts necessary for their description is not so very great: At any rate, it is limited.

During the past several decades, physics has grown and expanded to such an extent that workers in different branches cease to understand one another as soon as the discussion goes beyond what holds them together in one family, that is, beyond the limits of the laws and concepts underlying all branches of physics. Portions of physics are closely interwoven with technology, with other areas of natural science, with medicine, and even with the humanitarian sciences. It is easy to see why they have set themselve. up independent disciplines.

Surely no one would argue that any discussion of the various divisions of applied physics must be preceded by an examination of the basic laws and facts. And it is just as true that different writers select and arrange the material needed for laying the foundation of physics each in his own way, depending on his individual tastes and his own special field of inquiry What I have to offer here is merely one of many possible expositions of the fundamentals of physics.

The type of reader envisaged by this *Physics for Everyone* series has been mentioned in the prefaces to the earlier books. I will repeat that this series is aimed at representatives of all professions who wish to recall the physics they studied, get a picture of the modern state of the science, and evaluate the effect it has on scientific and technological progress and on forming a materialist world outlook. Many pages of these books will, I am sure, be of interest to teachers of physics and to students at school that have come to like physics. And finally there may be something of interest for those readers who are depressed by even a simple algebraic equation.

Quite naturally, this series is not intended to take the place of a textbook. *Photons and Nuclei* is an attempt, on the part of the author, to demonstrate to the reader how the laws of the electromagnetic field and quantum physics operate when we consider the behaviour of electromagnetic waves of different wavelength. Before taking up atomic nuclei, the reader will get some idea of what wave mechanics and the special theory of relativity are about. This is followed by a discussion of the basic facts concerning the structure of atomic nuclei, and then the topic will be sources of energy on the earth—a topic of burning interest to humanity at large. We conclude our brief talk with a story about the physics of the universe.

The limited scope of this book has forced us to give up many traditional topics. The old must always give way to the new.

A. I. Kitaigorodsky

CONTENTS

I. Soft Electromagnetic Radiation

Exchange of Energy by Radiation

Soft electromagnetic radiation is that with wavelengths lying roughly in the interval from 0.1 to 100 micrometres. Also, bear in mind that when we speak of soft radiation we mean electromagnetic waves not dealt with in radio engineering. This stipulation is important because purely radio-engineering methods permit one to dip into the region of soft radiation. The soft radiation also rather frequently goes by the simple term "light". When applying that term, one must bear in mind that visible light occupies only a very narrow section of wavelengths—for the "average" human eye it lies between 380 and 780 nanometres (or from 0.38 to 0.78 micrometres).

In future, whenever we have the occasion to speak of "light", we will do so in the broad sense of the word because the laws that hold true for the visible portion of the spectrum remain true for all other representatives of soft radiation.

Also note that radiation with shorter wavelengths than visible light is called *ultraviolet radiation*; the longer wavelengths are termed *infrared radiation.*

We can now turn to the topic of our discussion. It will be recalled that there are three modes of heat transfer. They are called heat conduction, thermal convection, and thermal radiation. In order to study the exchange of energy that occurs in thermal radiation, we will have to examine the behaviour of bodies in a vacuum (where convection is impossible) separated by a certain distance (this is to exclude the conduction of heat).

Experiments have shown that if two or more bodies form a closed system (the reader will recall that this means the absence of any exchange of energy between objects not in the system), the temperatures of the bodies equal out. Each one of the bodies of the system is at the same time a radiator and an absorber. What occur are numberless transitions of atoms and molecules from a higher level to a lower level (such events involve the emission of photons) and from a lower level to a higher level (photons are absorbed). Photons of all energies (or, what is the same thing, electromagnetic waves of all wavelengths) participate in these exchanges.

Quite naturally, the body does not absorb all the energy falling on it. There may be bodies that scatter more energy or transmit certain wavelengths. But this is of no consequence because sooner or later a thermal equilibrium is established nevertheless.

The condition of thermal equilibrium requires that the ratio of the energy of absorption to the energy of emission be the same for all wavelengths. This theorem was rigorously demonstrated in 1860 by the German physicist Gustav Robert Kirchhoff (1824-1887). The ratio can change for different temperatures, but if the temperature is fixed, then it is the same for photons of all energies.

This is a clear enough theorem and hardly needs any demonstration of proof. The idea behind the law is that the number of absorbed photons of a given kind (that is, of a definite energy) is equal, in the case of thermal equilibrium, to the number of radiated photons of that particular kind. From this we get the following rule: if an object is a strong absorber of any kind of rays, then those same rays are just as strongly radiated.

This rule helps to predict the conditions under which thermal equilibrium sets in. Why is water in a bottle with silvered sides so slow to heat up under the action of the sun's rays, whereas water in a flask made of black glass heats up very quickly? The explanation is obvious: a black body absorbs rays intensively and their energy goes to increase the temperature, and thermal equilibrium sets in after intense heating. A silvered surface, on the contrary, is an excellent reflector. Only a small amount of the energy is absorbed, it takes a long time to heat the body, and equilibrium sets in at a low temperature.

Now let's reverse the experiment. Pour some hot water into both flasks and put them into a refrigerator. Which one will cool off quickest? The one that heats up faster will cool off faster. If more energy is absorbed, more is released.

Some very effective experiments can be performed with coloured ceramics. If the object is green, the piece absorbs all colours except green. This is because the eye sees those rays that are reflected (or scattered) by the material. Now heat up the fragment. How will it appear? The answer is right at the tip of your tongue: violet because violet is the complementary colour of yellow-green. Complementary colours are those that produce white if they are mixed. Newton was the one who introduced the term "complementary colour" when he decomposed light rays into a spectrum with the aid of a glass prism.

The Radiation of Incandescent Bodies

It is well known that a piece of metal, when heated, first becomes red and then white. Most chemical substances cannot be thus heated. They either melt or decompose. Therefore, what follows refers mostly to metals.

The most remarkable thing is that the radiation spectrum of all heated bodies is not at all specific. The point is this. From the basic law about energy levels it is clear that the radiation spectrum and the absorption spectrum of a body must coincide. Metals are opaque throughout the region of the spectrum of soft radiation. From this it follows that they must also radiate photons of all energies.

Let's put this differently: a continuous spectrum appears due to the fact that in a multi-atomic system the energy levels of the atoms merge into overlapping bands. In such a system, all energy transitions are possible, that is, we can find any energy difference between the mth and nth levels $E_m - E_n$ and, hence, any frequencies of radiation and absorption. Figure 1.1 shows the spectrum of an incandescent body for several temperatures (we give here the theoretical curves that hold true for a so-called ideal black body).

It is well to point out here that the derivation of the shape of this curve (this was done by Max Planck in 1900) was the first step in the development of quantum physics. In order to obtain agreement of theory with experiment, Planck had to assume that radiation and absorption of light take place in separate portions. Planck was not prepared to take the next step and state that we are justified in speaking of particles of light (photons). That step was taken by Albert Einstein in 1905.

It was only in 1913 that Niels Bohr introduced the concept of quantization of energy. And if we want a logically rigorous theory of thermal radiation, the year of birth can be put as 1926.

Let's first discuss the shapes of the curves and only then talk about theory. First of all, note that as the temperature rises, the area under the curve rapidly increases. What is the physical meaning of the area bounded by the radiation curve? When constructing curves similar to those depicted in the figure, we say that the intensity of radiation for a given wavelength is laid off on the axis of ordinates (horizontal axis). But what does a "given wavelength" mean? Do we mean 453 or

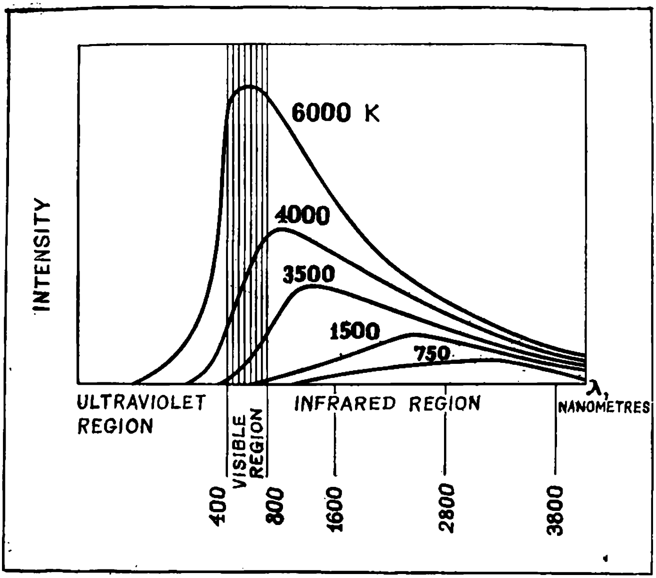


Figure 1.1

453.2 nanometres? Maybe 453.257859987654? It is probably clear that when we speak of a "given wavelength", we mean a very small interval of wavelengths. We agree, say, that the interval is equal to 0.01 nanometre. From this it follows that it is not the ordinate that has physical meaning but a tiny column with a base of 0.01 nanometre. The area of this column is equal to the energy radiated by the waves having lengths in that interval (for example, from 453.25 to 453.26 nanometres). Now if we break up the whole area into such columns, we get the total intensity of the whole spectrum. That is precisely the operation mathematicians perform and it is called integration. To summarize: the area under our curve yields the total

Max Planck (1858-1947)—outstanding German scientist who laid the foundations of quantum theory. In an attempt to find a mathematical expression for a proper description of the spectral distribution of the emission of an ideal black body Planck demonstrated that such a formula could be obtained by introducing a "quantum of action". Panck assumed that a body emits energy in parcels, equal to the product of a constant (which later was named after him) by the frequency of the light.

intensity of the radiation, and it turns out to be proportional to the fourth power of the temperature.

In the figure we are discussing it is clear that with increasing temperature there is not only a change in the area occupied by the curve but there is a shift of its maximum to the left, that is, into the region of ultraviolet radiation.

The relationship between the wavelength of light in micrometres that corresponds to the greatest intensity of radiation (or absorption) and the temperature in kelvins is given by the following formula:

$$\lambda_{max} = \frac{2886}{T}$$

At the lowest temperatures, the maximum lies in the infrared region. That is precisely why infrared radiation is also termed thermal radiation. And it is a marvelous thing that we have instruments capable of sensing the thermal radiation emitted by bodies at room temperature and even lower. There are instruments today that can see in total darkness. Some animals, by the way, have the same capability. There is nothing at all strange in this fact since infrared rays have, in principle, the same properties as visible rays.

Also, don't forget that every animal is a source of radiation. We sometimes hear of a person being able to "feel" the presence of another person in darkness. No mysticism is involved, merely the one who "feels" has a highly sensitive perception of thermal rays.

I can't resist telling the reader about an interesting episode that demonstrates the necessity of taking thermal rays into account even when, in the very ordinary sense of this word, the source of rays is not a heated body. A few years ago I was asked to investigate some experiments conducted by a person who considered himself a magician capable of stopping a motor using only his will power.

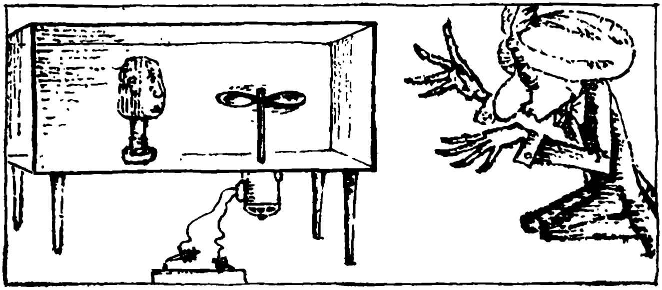


Figure 1.2

My task was to find a rational explanation (sorcerers of the twentieth like to deal in pseudoscientific terminology and so call these experiments telekinesis).

A diagram of the experiment is shown in Figure 1.2. A wing was set in rotation by a small motor, and the wing really did stop whenever the magician sat down next to the box where the axle of the motor emerged from below. I soon found that anybody who sat down in that position would be able to stop the wing. It usually took about 10 to 15 minutes to perform the operation. And it wasn't the motor that came to a halt, as the magician claimed, but the little wing. It was clear then that some kind of force connected with the human body was interfering with the force of adhesion between the axle of the motor and the wing.

I pointed out that the wing could be stopped almost instantly if an electric lamp were brought up close to the side of the box. It was obvious that the trick lay in the heat emitted by the human body. I sent a whiff of tobacco smoke into the box and demonstrated that the convection currents of air inside the box move so as to prevent the wing from rotating. Exact measurements showed that the temperature of the side of the box closest to the human body is about one degree higher than the opposite side.

Infrared rays emitted by a body heated to 60-70 degrees Celsius can be felt if the hand is held at a short distance. Thermal convection of course must be eliminated. Heated air rises upwards, so bring your hand close from below. You will then be certain that you are perceiving thermal rays.

We conclude our talk about thermal rays with an explanation of why the modern electric light bulb with a tungsten filament is a great step beyond a bulb with a carbon filament. The whole point is that a carbon filament can be heated to 2100 K, while a tungsten filament can be heated all the way up to 2500 K. Why are these 400 degrees so important? Because the purpose of an incandescent lamp is to provide light and not heat, and so the aim is to have the maximum of the curve located in the visible portion of the radiation. It will be seen from the graph that the ideal is to have a filament that can withstand the temperature of the sun's surface, or 6000 K. But even the step from 2100 degrees to 2500 degrees raises the portion of energy involving visible radiation from 0.5% to 1.6%.

The Theory of Thermal Radiation

If a system of radiating and absorbing bodies is closed, then the photon "gas" (with the aid of which the bodies exchange energy) must be in equilibrium with the atoms supplying the photons. The number of photons with energy $h\nu$ depends on how many atoms lie in the E_1 level and how many in the E_2 level. In the case of equilibrium, these numbers remain unchanged.

However, the equilibrium is of a dynamic nature since the processes of excitation and radiation occur at the same time. In some way (either by collision with another particle or due to absorption of a photon from without) the atom or the atomic system climbs to a high level. The system persists in this excited state for some (indefinite) time (ordinarily for a fraction of a second) and then reverts to a low level. This process is termed *spontaneous radiation*. The atom behaves like a little ball on the top of a sharp peak with an intricate configuration: the slightest breath of air is enough to disrupt the equilibrium. The ball rolls down into a valley, usually the lowest part, and then only a strong impact can bring it out again. We say that an atom that has dropped to the lowest level is in a stable state.

But here we must bear in mind that there are also intermediate states in between the peak and the lowest portion of the valley. The ball may be at rest in a slight depression from which it can be extricated by a waft of air, so to speak, or at least by a little push. This is a metastable state. Thus, besides the excited and stable states there is a third, metastable, type of energy level.

To summarize, then, the transitions will occur in both directions. First one atom and then another atom will move into a higher level. In the next instant, they will fall to a lower level and emit light. But at the very same time, other atoms will receive energy and will rise to upper levels.

The law of conservation of energy requires that the number of transitions upwards equal the number of transitions downwards. What does the number of transitions upwards depend on? Two factors: first, the number of atoms in the lowest floor, and, second, the number of impacts that raise them to a higher floor. And the number downwards? It is of course determined by the number of atoms lying in the upper floor, and it would seem to be independent of any other factors. That is precisely what theoretical physicists thought at first, and yet the pieces didn't fit. The number of transitions upwards, which is dependent on two factors, increased with temperature much faster than the number of transitions downwards, which is dependent on only one factor. This model, which had appeared to be so obvious, turned out to be nonsense: Sooner or later all the atoms would be chased up to the highest level: the system of atoms would be in an unstable state with no radiation.

It was precisely this impossible conclusion that Einstein, in 1926, picked up from the reasoning of his predecessors. Apparently, there was some other influence affecting the transitions of atoms from the upper floor to the lower floor. One could only conclude that there is a forced transition in addition to the spontaneous transition to the lower level.

What is this *stimulated emission*, as it is called? In short, it is this. A system is in the upper level. It is separated from the lower level by the difference $E_2 - E_1 = h\nu$. Now, if a photon with energy $h\nu$ is incident on the system, then it makes the system move down to a lower level. The incident photon is not absorbed in the process but continues onwards accompanied by a fresh photon of exactly the same kind generated by the first one.

Do not seek any logic in this reasoning. It was intuition, a guess and experiment was to prove it right or wrong. Using the assumption of stimulated emission we are able to derive a quantitative formula that yields the graph of emission as a function of the wavelength of a heated body. The theory proved to be in brilliant agreement with experiment and so justified the hypothesis.

It is an exciting thought that the practical conclusions from the fact of the existence of stimulated emission that led to the invention of lasers were drawn only many years later.

Optical Spectra

Generally speaking, any body is a source of soft electromagnetic radiation. Using a spectrograph (this is an instrument whose principal component is a prism or a diffraction grating), light can be decomposed into a spectrum. The spectrum may turn out to be continuous, banded or line. The spectra of incandescent solids are very similar. For that matter, only a few substances can be heated to incandescence. A real rarity is a glowing liquid. The emission spectra of gases are highly informative. Such are the spectra of rays coming to us from distant stars. The bulk of the information we have concerning the structure of the universe comes to earth in the form of light rays of stellar matter in a gaseous state.

Under terrestrial conditions, it is easy to obtain emission spectra of atoms. Atoms are made to glow either by passing a current through the gas or by heating. In this manner we can obtain only the spectra of atoms but not the spectra of molecules. Before the gas begins to glow, the molecules break up into atoms. That is why absorption spectra are studied if the investigator is interested in liquids or solids. In the final analysis, the picture is determined by the system of energy levels. Transitions up or down yield the same information. Simply, do what is easiest.

Spectra that consist of separate clear-cut lines can be obtained only from a gas or a diluted solution. In the second book of this series it was stated that the behaviour of dissolved molecules resembles in many respects the behaviour of a gas. This also holds true for optical spectroscopy. Unfortunately, the solvent affects the character of the spectrum, but if we compare the type of spectra of molecules dissolved in different substances, it is possible to take that effect into account and extract from the experiment the fingerprints of the dissolved molecule.

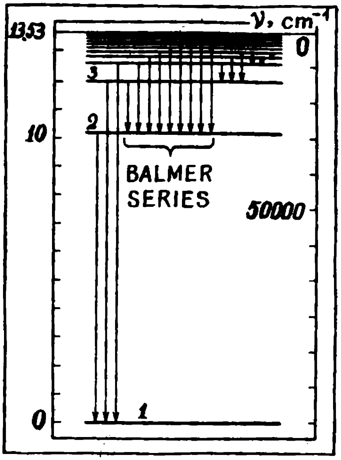


Figure 1.3

Obtaining a characteristic spectrum does not mean establishing the system of energy levels of a molecule. But for many practical purposes this is not required. With an album of information about spectra (that is, the list of spectral lines and their intensities, or the curves of intensity versus frequency) of some family of chemical substances, we can, by taking the spectrum of an unknown substance and comparing the experimental pattern with material from the album, determine the substance in the very same way that a criminal is detected from the fingerprints he leaves.

Just lately, optical spectral analysis has come up against a competitor: radiospectroscopy. Radio spectroscopic methods are still inferior in sensitivity to optical methods (though the inferiority will most likely not last long) but are far superior to optical methods in the identification and quantitative analysis of mixtures of substances.

We don't aim here to acquaint the reader with concrete spectra of substances. It will suffice to discuss the pattern

Niels Bohr (1885-1962)— the famous Danish physicist who created the first quantum model of the atom and thus discovered the law of quantization of energy. He was an active participant in developing the principles of quantum mechanics. He demonstrated the fundamental inapplicability — to the microworld — of concepts suitable in describing the behaviour of macroscopic bodies. He made a very considerable contribution to the theory of the structure of the atomic nucleus.

of energy levels of atoms of hydrogen and the fundamental scheme of energy levels of a free molecule.

Figure 1.3 depicts the system of energy levels of hydrogen. Note the characteristic thickening of levels as we move away from the zero line.

Incidentally, the zero in the diagram is not a "real" zero actually. An unexcited atom of hydrogen naturally possesses some energy. But since spectra exhibit energy differences, it is convenient to reckon energies from the lower line. Depending on the intensity of the "shock" obtained, the atom can rise to any one of the "floors", hold on for a moment in the nonequilibrium state and then, via one of two possible modes (spontaneous emission or stimulated emission), revert to the lower level.

The resulting spectrum may conveniently be split up into a number of series. Each series is subordinate to its lower level. In the visible portion of the spectrum we have the so-called *Balmer series*. The explanation of this series was the first triumph of Niels Bohr's theory of atomic structure.

Not all energy transitions are of equal probability. The higher the probability of transition, the more intense the appropriate line. There are also forbidden transitions.

A great achievement of theoretical physics was the exhaustive interpretation of the spectrum of the hydrogen atom via the solution of the famous equation of quantum mechanics derived in 1926 by the Austrian physicist Erwin Schrödinger (1887-1961).

Atomic spectra are affected by external fields. The lines split into several components under the action of an electric field (the *Stark effect*) and under the action of a magnetic field (the *Zeeman effect*). We will not go into these exciting phenomena here, but it is necessary to point out that an understanding of them came only after Samuel Goudsmit and George Uhlenbeck made the assumption that the electron possesses spin. How spin

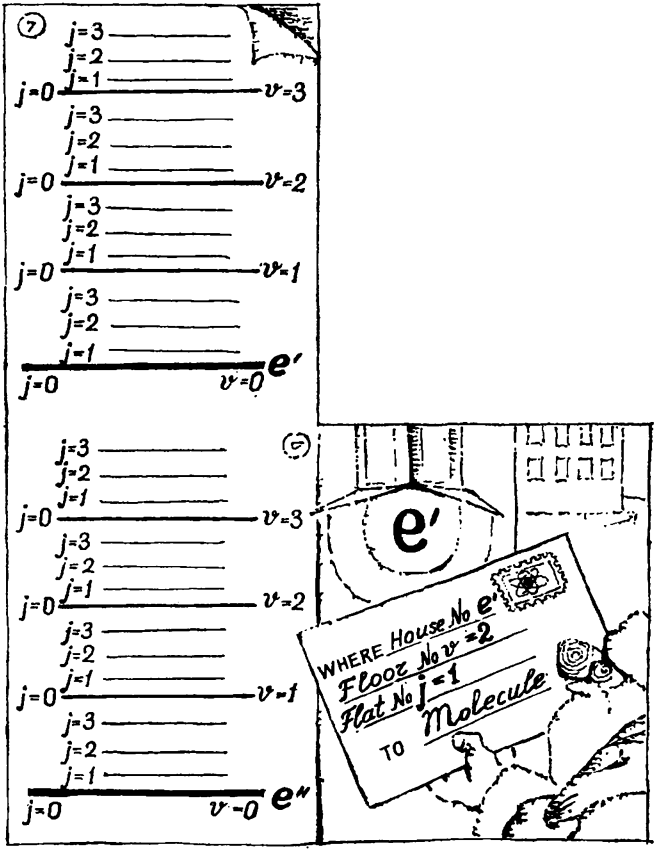


Figure 1.4

reveals itself in experiment directly was discussed in the third book of this series.

And finally the last remark regarding the pattern of energy levels. We see that the limit which the levels come up to is marked by the number 13.53. What kind of number is this? This is the *ionization potential*. If we multiply the electron charge by the magnitude of this potential in volts, we obtain the amount of work needed to tear the electron away from the nucleus; in other words, this is the work that must be done to destroy the hydrogen atom.

Atomic spectra arise as a result of electron transitions. As soon as we move from atoms to a molecule, we immediately have to take into account two more components of energy. A molecule can rotate and the atoms of a molecule can perform vibrations with respect to one another. All these types of energy can likewise be quantized, which means they can have only specific discrete values. Thus, the energy state of a molecule is described by the state of its electron cloud (*electronic level*), the state of oscillatory motion (*vibrational level*), and the state of rotation (*rotational level*). We thus have to deal with three kinds of information: the number of the house, the floor, and the flat.

But what is the role of the "floor" and the "flat"? What energy levels have big separations and what levels have small separations? All these questions are answered in Figure 1.4. This diagram shows two electronic levels e' and e'' (the house numbers). The floors are the vibrational levels marked v, and the numbers of the flats are the rotational levels marked j. True, this differs somewhat from the ordinary numbering of houses and flats, which is continuous; in dealing with molecular spectra we number the flats on every floor from zero. Thus we see that the gaps between rotational levels are the smallest and between electronic levels (e' and e'') the largest.

Suppose a molecule has the following possible electronic levels: 100, 200, 300, units of energy, vibrational levels at 10, 20, 30, ... units of energy, and rotational levels at 1, 2, 3, ... units of energy; then a molecule on the second electronic level, the first vibrational level, and the third rotational level will have an energy of 213 units.

Thus, we can give the energy of a molecule as follows:

$$W = W_{el} + W_{vib} + W_{rot}$$

The frequency of the emitted or absorbed light will always correspond to the difference (symbol: Δ) of two levels:

$$\nu = \frac{1}{h}(\Delta W_{el} + \Delta W_{vib} + \Delta W_{rot})$$

I would like to touch on those transitions that involve a change in only one type of energy. In practice, this occurs only in the case of rotational transitions, and why this is so will soon be seen.

We begin our investigation with the absorption of electromagnetic waves of a group of molecules starting with the longest wavelengths, that is, with the smallest portions of energy $h\nu$. Until the magnitude of the energy quantum has become equal to the distance between the two closest-lying levels, the molecule will not begin to absorb. By gradually increasing the frequency, we will reach quanta capable of raising the molecule from one "rotational" step to the next. Experiment shows that this occurs in the region of microwaves (the edge of the radio spectrum) or, to put it differently, in the region of the far infrared spectrum. Wavelengths of the order of 0.1 to 1 millimetre will be absorbed by the molecules. What we then have is a purely band spectrum.

New things happen when we irradiate the substance with energy quanta sufficiently high to move the molecule from one vibrational level to another. However, we will never attain a purely vibrational spectrum, that is, a series of transitions under which the number of the rotational level is preserved. On the contrary, transitions from one vibrational level to another will involve a variety of rotational levels. Say, a transition from the zero (lowest) vibrational level to the first can consist in moving up from the third rotational level to the second, or from the second to the first, and so forth. We thus obtain a vibration-rotation spectrum. We observe it in infrared light (from 3 to 50 micrometres). All transitions from one vibrational level to another will differ slightly in energy and will yield a group of very close lines in the spectrum. In the case of small resolution, these lines merge into a single band. Each band corresponds to a definite vibrational transition.

We thus move into a new spectral region, the region of visible light where the energy of the quantum becomes sufficient to move the molecule from one electronic level to another. Of course, it is not possible here to obtain either purely electron transitions or electron-vibrational transitions. Complex transitions arise in which the energy transition is accompanied by a change in the "house", the "floor", and the "flat". Since a vibrational-rotational transition constitutes a band, the spectrum in the visible region will be practically continuous.

The characteristic spectra of atoms and molecules have for many years played the role of helpers in determining the chemical structure and composition of substances. And their aid continues today. Revolutionary events in the field of spectroscopy have only just recently occurred.

Laser Radiation

The first thirty years of this century saw fantastic advances in theoretical physics with the discovery of such important laws of nature as the laws of the mechanics of high velocities, the laws of the structure of the atomic nucleus, and the laws of quantum mechanics. And the forty years that followed exhibited just as phenomenal a development in the applications of theory to practice.

This was a time when humanity harnessed the energy of atomic nuclei and invented semiconductor transistors that revolutionized radio engineering and led to the development of electronic computers and laser technology. These three applications were actually what produced the modern revolution in science and engineering.

In this section we discuss lasers. First let us give some thought to the problem of why, operating via traditional methods, we are not able to generate an intense directed beam of light.

The strongest stream of light collected into a very narrow beam disperses and loses its energy over small distances. It was only in the science fiction of the Russian writer Aleksei Tolstoi that the hero devises a "hyperboloid" capable of generating bursts of intense light rays that can burn and cut materials and carry tremendous energy over great distances. Of course we know that it is possible to manufacture a concave mirror that can generate a parallel beam of light. This requires placing a point source in the focus of the mirror. But a point is a mathematical abstraction. All right, suppose we have a small source, bigger than a point. But even then, if we heat the ball to 6000 degrees (and no known material can stand more), we obtain a beam of light of miserably low intensity. And as soon as we increase the dimensions of the source, then instead of a parallel beam of rays we obtain a spread-out fan of light "filaments" and the intensity of the ray of the projector begins to rapidly diminish with distance.

Thus, the first obstacle to creating a strong beam of light is that atoms emit light in all directions. That's the first, but not the last. Atoms and molecules emit light without agreeing on how they'll do it. The result is that rays from different atoms set out at different times, totally unmatched in their efforts. The emissions of different atoms do not agree in phase, and that means that rays from different atoms will frequently annihilate one another. This occurs, as you will recall, when the hump of one wave comes with the valley of another one.

Laser emission is what overcomes these obstacles. The word "laser" stands for *l*ight *a*mplification by *s*timulated *e*mission of *r*adiation.

The underlying idea is made up of several elements. First of all, recall that there is stimulated emission and spontaneous radiation. We have already mentioned that this type of emission occurs when a light photon encounters an excited atom. If the excitation energy of the atom is exactly equal to the energy of the photon, then the photon de-excites the atom, which moves to a lower level and emits a photon. The marvelous peculiarity of stimulated emission is that the new photon is the same as the one that generated it, and not only in energy but also in phase and direction of motion.

The second element behind this idea is this. If the system of emitting atoms is placed in a tube, the ends of which are at a certain distance from each other and can serve as mirrors for the photons that interest us, then we can build up a bunch of photons generated (as they move back and forth) by identically excited atoms.

The third element in this idea is to retain the atoms in the excited state as long as possible and then, after the pumping is completed, to force all atoms to de-excite at the same time. Putting this laser idea (that is, the production of millions upon millions of identical photons from a single photon) into hardware should permit generating a light beam of unimaginable intensity. Such a beam would exhibit only the slightest spread and would have tremendous energy over its cross section.

But the question is: How can this be attained? For decades no one knew. Back in the 1930s, important ideas in this connection were expressed by Soviet physicist V. A. Fabrikant. Later, persistent research of Soviet

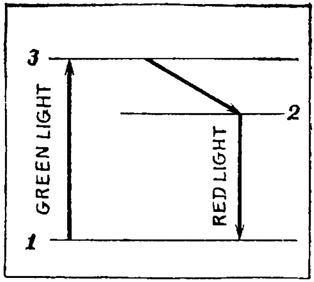


Figure 1.5

scientists A. M. Prokhorov and N. G. Basov and, independently, the American physicist Charles Hard Townes led to the invention of lasers. All three received the Nobel prize in physics.

Suppose a system has two energy levels. Most atoms or molecules are in the lower level. Thermal shocks can transfer a molecule to the upper level for a short time. But not for long, because the molecule is de-excited. In the process most of atoms go to the lower level spontaneously. Of course, some of emitted photons will carry some of the excited atoms to the lower state and generate photons of stimulated emission. But these are rare processes because there are few excited particles (the most occupied are the lower levels), and also the probability of a spontaneous transition is substantially higher than the probability of stimulated emission.

Let us suppose it was possible to find a substance whose atoms have the three energy levels marked in Figure 1.5 by the numerals 1, 2, and 3. The distance 1-3 corresponds to the frequency of emission of green light, the distance 1-2 corresponds to the frequency of red light. Now suppose the probability of a transition from level 3 to level 2 is thousands of times higher than the frequency of transitions from level 2 to level 1. Let us irradiate the substance with green light. The atoms will rise to the third floor, then via spontaneous transitions will go to level 2 and will stay at that level. This is termed *nonradiative transition*. The energy released goes into the vibrational energy of the atoms. Using our imagination further, let us suppose that we have carried most of the atoms to level 2. We have thus reversed the occupancy density, it is no longer "normal". There are more in the upper levels 2 than in the lower levels 1, which is impossible when the process is controlled solely by thermal motion.

And still and all there does begin a transition from level 2 to the lower level 1. An appropriate photon will encounter other atoms in the excited level 2. The result will be not absorption but the creation of a new photon. The first, accidentally generated photon in 2-1 will be joined by the very same photons of stimulated emission.

Thus arises a stream of 2-1 photons. They will all be identical and will generate a beam of tremendous intensity.

That precisely was the process that the three Nobel prize winners were able to create. Historically, the first was a ruby laser. The diagram of levels shown in the figure is precisely the diagram of ruby with an admixture of chromium atoms.

To make a laser we need a source of excitation that does the pumping of the laser, that is, that carries the atoms to higher levels.

If the source of laser emission is a solid, then it is made in the form of a cylinder whose bases play the part of mirrors. In the case of a liquid or gaseous laser, a tube is constructed with mirrors at the ends of the column. By performing a micrometre-precise positioning of the mirrors and thus fixing the length of the column, we put in a privileged position only those photons whose integer number of wavelengths fit into the length of the column. Only then do all the waves combine.

Perhaps the main peculiarity of the laser is the possibility of creating a narrow stream of radiation. Practically speaking, a laser beam can have almost any cross section. Technically, this is achieved by the fact that the ray is made to travel along a narrow glass capillary tube of sufficient length. Photons moving at an angle to the capillary do not participate in the photon build-up. A resonance cavity (that is, the mirrors that reflect photons first in one direction and then in the other during the pumping period in the operation of the laser) reproduces photons of only one direction. In some cases, if an angular dispersion of the beam of the order of one degree is not satisfactory, a supplementary lens is placed in the path of the released ray.

A laser device is a complicated piece of engineering if one has to do with high power outputs. A primary pulse is first set up in the column; then it is fed to amplifiers that function in the same manner as the first column but pump independently of the first column. We will not go into these details because we are only interested in the physical principles of pumping and the generation of laser emission. They can differ greatly, as is evident from a glance at figures 1.6 to 1.8 with diagrams of the action of lasers which today yield beams of maximum power output.

Figure 1.6 depicts a so-called neodymium laser. Actually, the body of the laser is not the metal neodymium but ordinary glass with an admixture of neodymium. Ions of neodymium atoms are haphazardly distributed among the atoms of silicon and oxygen. Pumping is performed by flash bulbs. The lamps emit in wavelengths between 0.5 and 0.9 micrometre, and a broad band of excited states is obtained (shown here in the form of five bars). The atoms perform nonradiative transitions to the upper laser level (labelled 2 in all three figures). Each transition yields a different energy, which is converted into the vibrational energy of the whole "lattice" of atoms.

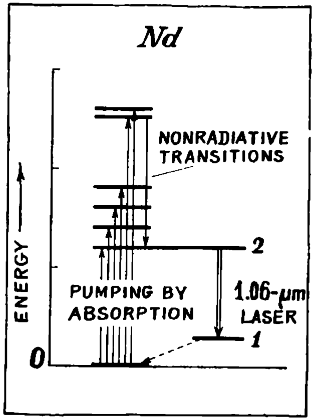


Figure 1.6

Laser emission, that is, the transition to an empty lower level labelled 1, has a wavelength of 1.06 micrometres.

The dashed line, which depicts the transition from level 1 to the lowest level "does not work" (in the sense that energy is released in the form of noncoherent radiation).

A neodymium laser permits obtaining fantastic power outputs of up to 10^{12} watts. The energy is generated in the form of pulses lasting 0.1 nanosecond.

A new competitor in this field is a laser using transitions in excited atoms of iodine (Figure 1.7). The working substance here is the gas C_3F_7I. Here, too, flash bulbs are used for pumping, but the physical processes are different. Ultraviolet light of wavelength 0.25 micrometre is used for pumping. Under the action of this radiation there occurs a dissociation of the molecules. The remarkable thing is that the iodine atoms are torn out of the molecule and are in an excited state! As the reader will see, this is quite a different method for inverting the

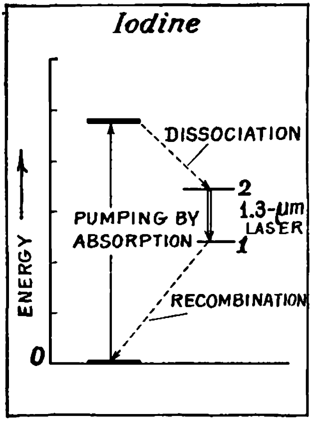


Figure 1.7

occupancy density. The operating transition 2→1 leads to laser emission with a wavelength of 1.3 micrometres, after which the iodine atom joins up with the molecular residue.

The reader has also probably heard of the widespread use of helium-neon lasers. They are used to obtain an intense infrared ray of wavelength 1.13 micrometres. These lasers are not record holders as far as power goes, and so we give a diagram of levels for a different laser that operates on a mixture of nitrogen and carbon dioxide (Figure 1.8).

Before describing this laser, a natural question comes to mind, and that is: Why are mixtures of gases needed? The point is that some atoms and molecules are more easily excited while others are more easily de-excited, so that in a laser operating on a mixture, particles of one type effect the pumping process, these then transfer the energy via collisions to other atoms or molecules, and these in turn generate the laser beam. There are systems now functioning that consist of more than two gases.

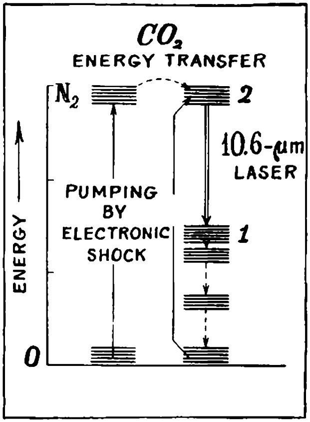


Figure 1.8

For instance, in the nitrogen-carbon dioxide laser, it is advisable to add a variety of components including helium.

Pumping in the CO_2 laser is done differently from the two just described. The mixture of gases is placed in a gas-discharge tube and a sufficiently high voltage is applied so that the system becomes a plasma. Electrons moving at high speeds excite the vibrations of nitrogen molecules. The diagram shows a transition of this molecule to the upper floor. The voltage applied to the electrodes plays a delicate role. The optimal energy for exciting nitrogen molecules is about 2 eV

The nitrogen molecule is only an intermediary. It does not by itself produce any emission but rather transfers the energy obtained from the electrons to the CO_2 molecule and lifts it to an upper laser level.

The upper laser levels 2 are "flats of the third floor" of CO_2 molecules. A molecule of gas has a lifetime in the upper laser level equal to about 0.001 second. This is not so little, and the molecule has a good chance of encountering a photon of suitable energy that will force it down to a lower level.

It should be pointed out that "interflat" transitions are a much more frequent occurrence than "interfloor" transitions. Lifetimes on the rotational level are of the order of ten millionths of a second. This favourable circumstance results in the flats of each floor being occupied in a rather stable fashion, and so, using an engineering technique that we have already mentioned (setting a suitable distance between the mirrors), it is possible to isolate some one transition; let us say, from the sixth flat of the third floor to the fifth flat of the second floor.

The designing engineer must have at his disposal complete information about the residence time of an atom on any given sublevel and about the probabilities of transition. Then he is able to choose the optimal radiation of the given gaseous mixture. Ordinarily, a laser operating on carbon dioxide is tuned to a wavelength of 10.5915 micrometres. For a laser to function normally, it is necessary that the molecules not pile up on the lowest laser level, the idea being for them to do their jobs and get out. Now, at a gas pressure of 1 millimetre of mercury, carbon dioxide molecules experience 100 collisions per second in vacating the level. The respective figures are 4000 and 100 000 if helium and water are present. This is a tremendous difference.

By selecting the proper admixtures for carbon dioxide, we can boost the power output of the instrument substantially. It would appear that this is the gold-medal winner in the laser field.

A carbon dioxide (CO_2) laser produces a beam that can be focussed on an area of 0.001 cm^2 with an intensity of 1000 kW/cm^2 in continuous operation and one million kW/cm^2 in pulsed operation with the pulse time equal to one nanosecond (which, as you know, is 10^{-9}, or one thousand millionth, of a second).

The search for suitable materials for lasers is a sort of art. One needs intuition, ingenuity, and memory to create an effectively operating laser. The user can now order lasers with a great variety of wavelengths ranging from a tenth of a micrometre to hundreds of micrometres.

The exceptional intensity and coherence of laser emission have revolutionized many areas of engineering. During the past decade, the manufacture of lasers has become a whole industry. Lasers have found application as generators of radiation that trasmit not only energy but also information. Intense research is in progress for their use in initiating thermonuclear reactions. Lasers are used in place of the surgeon's knife in medicine, as an instrument for the most delicate surgical operations, as a device for the separation of isotopes. We will have occasion later on to come back to further discussions of the marvelous laser.

Luminiscence

Thermal radiation is a universal property of all bodies. Thermal rays are emitted by every body at any temperature from absolute zero upwards. The thermal spectrum is a continuous one and is depicted by a curve that we have already discussed. True, our curve was that of a black body, but the behaviour of coloured bodies is in principle but slightly different from the behaviour of black bodies. Merely, the curve for coloured bodies is distorted somewhat. But the general increase in the energy of emission (as the temperature rises) and the displacement of the maximum to the left (if wavelengths are laid off on the axis of abscissas) are the general law.

All radiation consists in a transition from a higher energy level to a lower level. But the reasons for excitation of atoms or molecules may differ. In the case of thermal radiation, it is the collisions of particles of the substance due to thermal motion.

But that is not the only reason compelling a body to emit waves. *Luminescence*, which we are about to discuss, is of a different nature. This term embraces processes of excitation of molecules that are not connected with any increase in the temperature of the body. Causes of particle excitation may be encounters with beams of photons or electrons, mechanical impact, friction, and so on.

Practically all substances are capable of luminescence. But only some (called luminophors or phosphors) glow brightly and are of practical importance.

Luminophors are used as materials for covering television and oscillograph screens, in which case the luminescence occurs under the impact of electrons. Certain substances luminesce brightly under the action of ultraviolet radiation. The energy of the incident photon must be at least greater than the energy of the emitted photon. That is why the incident quantum of energy can come from the invisible portion of the spectrum while the emitted radiation can lie in the visible portion.

Admixtures of luminescent material measured in minute fractions (billionths, or 10^{-9}) are sufficient to make a substance luminesce under irradiation by ultraviolet light. That is why fluorometric analysis is sometimes used as a tool in chemical analysis. It is capable of detecting minute quantities of impurities.

Luminophors are used to cover the walls of daylight lamps.

There are two types of luminescence: fluorescence and phosphorescence. *Fluorescence* consists in the de-excitation of an atom or molecule that occurs without the molecule remaining in the excited level. Contrariwise, *phosphorescence* persists after the excitation has ceased. This occurs if the system, when excited, passes to a metastable level, from which transitions downwards have a low probability. As a rule, the radiation occurs after the molecule first absorbs the energy and rises to a higher level, after which de-excitation takes place, the transition to the lower level occurring without any stop at an intermediate, metastable, level.

A few words about *electroluminescence* that occurs in certain semiconductor diodes on the boundary of a *p-n* region. This interesting phenomenon is of great practical value because it underlies the manufacture of semiconductor lasers. The idea is this: an electron and hole of the semiconductor can recombine with the emission of a photon.

For transitions of this type to take place continuously we have to pass an electric current through the diode. The problem is to find a suitable material that satisfies several requirements. First of all, the current has to inject (if that is the word) electrons into the *p*-type semiconductor material, that is, a semiconductor containing more holes, or it must pump holes into an *n*-type crystal. This is a necessary condition, but other factors such as, for example, the rate of transition from an upper to a lower level can play a decisive role. Then there are cases where all factors favour a transition of an electron downwards and electroluminescence takes place.

A particularly good electroluminescent material is the semiconductor gallium arsenide. It yields a sufficient quantity of photons. The photons move along the *p-n* boundary. Two sections of the diode perpendicular to the boundary are polished and that sets up a resonant cavity. The photons generated in recombinations of holes and electrons are in phase, and for sufficiently large currents the radiation becomes stimulated emission with all the resultant consequences of being narrow, highly directed, and polarized.

Semiconductor lasers operate in a band of wavelengths from the ultraviolet to the far infrared and are widely used for a variety of purposes.

2. Optical Instruments

The Prism

The armamentarium of instruments employed in laboratories and industry have such a big turnover that if a scientific worker dropped out of research for a decade or two he would have to start studying all over again. But today and, most likely, in the distant future he will again meet his old acquaintances, the prism and the lens. Let us recall the simple laws that light obeys in interactions with these transparent materials. Incidentally, transparency is a relative notion. For certain electromagnetic waves, even wood and concrete are transparent.

The laws of interaction of a light ray and a body capable of reflecting and refracting the ray are simple until the wave aspect of the light waves becomes involved. These are the *law of reflection* (the angle of incidence is equal to the angle of reflection) and the *law of refraction* of light. It will be recalled that when a light ray falls on the boundary between two media, it is deflected from its rectilinear path. The angles of incidence i and of refraction r are connected by the relation

$$n = \frac{\sin i}{\sin r}$$

This law was established, in very careful measurements, by the Dutch physicist Willebrod Snellius or Snell (1591-1626), professor at the University of Leiden. The contents of his course of lectures in which he described the phenomena of light interacting with transparent bodies was well known to a small (in those days) circle of European scholars.

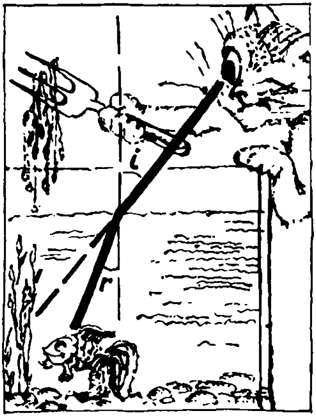


Figure 2.1

That most likely was the reason René Descartes' (1596-1650) article entitled *Discours de la Méthode* (1637) was scoffed at by his contemporaries, for it was here that Descartes would appear to have "proved" this law with the aid of rather strange reasoning. Descartes' nebulous verbiage did not excite the admiration of his colleagues. But the fact that his discourse did indeed result in the correct formula was explained away very simply: he was said to have fitted arguments to a result that had already been obtained. And so Descartes had also to experience accusations of plagiarism.

Perhaps, after all, we might join the sceptical attitude of his contemporaries. Descartes considers a ball thrown onto a weak net. The ball breaks through the net and loses half its speed. Then, writes the great philosopher, the motion of the ball differs radically from its original design, in one direction or another. It is indeed hard to get at what the real meaning is. Perhaps Descartes wanted to say that the horizontal component of velocity of the ball does not change while the vertical component does, for it is precisely in that direction that the net obstructs the motion of the ball.

But let us get back to the law of refraction.

The angles i and r are ordinarily laid off from the position of the normal as shown in Figure 2.1. The quantity n, which is called the index of refraction (or refractive index), depends on the media involved. In order to compare bodies as to optical properties, it is convenient to set up a table of the indexes of refraction for the case of an incident ray from the air (or, if one is a pedant, from a vacuum) into a medium. In this case, the angle of refraction is always less than the angle of incidence, and hence the refractive index is greater than unity.

The refractive index grows with the density of the medium. For diamond, it is 2.4 and, for ice, it is 1.3.

I'm not going to give a table of refractive indexes, but if I did, I'd have to indicate for what wavelength of light the data are given. The index of refraction depends on the wavelength. This is an important phenomenon underlying operation of a number of instruments that resolve electromagnetic radiation into a spectrum and is called dispersion.

If light falls from one medium into a medium of smaller density, then a complete internal reflection can occur. In this case, the refractive index is less than unity. As the angle of incidence increases, the angle of refraction will approach 90° Provided that $\sin r = 1$, $\sin i = n$ the light will cease to pass into the second medium and will be reflected *in toto* at the interface. For water, the angle of total internal reflection is equal to 49°

The refraction of light by means of a plane plate can be used to shift a ray to a position parallel to itself. And with the aid of a prism a light ray can even be turned around.

If the reader wants to recall the derivation of the formula for the angle of rotation D of a light ray, he can find it in a school textbook. The derivation only requires a few facts from elementary geometry, but it is very unwieldy, particularly if it is done for a thick prism and for an arbitrary value of the angle of incidence of the ray on the prism. A simple formula is obtained if the prism is thin, and the angle of incidence of the ray on the face of the prism does not differ greatly from a right angle. If that is the case, then

$$D=(n-1)\,p$$

where p is the angle between the faces of the prism.

Using a prism, the great Newton demonstrated for the first time (this was at the end of the seventeenth century) that white light is not monochromatic but consists of rays of different colours. Violet rays undergo the greatest deflection, and red the smallest. That is precisely why we say ultraviolet and infrared rays, and not infraviolet and ultrared.

The scientific world learned of Newton's discovery in 1672. In explaining his experiments, Newton is clear and exact. Therein lies his genius. As for his discussions of the matter, it is no easy job to plough through them. Only after much digging in the verbiage can one gather that although the author had promised to depict facts and not to create hypotheses (Newton's famous phrase: "hypothesis non fingo", or "I do not frame hypotheses") he did not carry out his promise. Many of the axioms and definitions, like, say, "a ray of light is its minutest part" are very strange indeed to the modern ear.

In chemistry the spectrograph still reigns supreme, and the main component is Newton's prism. The material must possess a high degree of dispersion. Prisms for spectrographs are made out of quartz, fluorite, and rock salt. The light to be resolved is passed through a slit located in the principal focal plane of the input lens. That is why a parallel beam of light falls on the prism. Photons of different frequency go in different directions. The second, or exit, lens collects identical photons into a single point of the focal plane. The spectrum may be viewed with the naked eye, but then a piece of frosted glass is required. The spectrum can also be photographed.

At the present time, spectra are registered by automatic recorders. An energy receiver in the form of a photocell or thermocouple slides along the spectrum. The receiver generates a current whose strength is proportional to the light intensity. This current deflects the moving part of the recorder in exactly the same way that the current of a galvanometer deflects the needle. The deflected part has an attached stylus that records the spectrum on a roll of paper tape that unwinds at a constant rate.

The Lens

A whole industry is engaged in the manufacture of lenses. These are transparent bodies bounded by two spherical surfaces or one spherical surface and one plane surface, and they come in all imaginable sizes. Some devices use lenses the size of a small coin, in others (large telescopes) there are lenses several metres across. The manufacturing of large lenses is a real art, because a good lens must be homogeneous throughout.

Every reader knows what a lens is and probably knows the main properties of one. A lens magnifies, a lens can focus rays of light. Using a lens placed in a strong beam of light (from the sun), it is easy to set a piece of paper afire. A lens "collects" rays of light in a single point, the focus (focal point) of the lens.

The fact that parallel rays converge to a single point and, conversely, that a lens produces a parallel beam of rays if a point source of light is placed in the focus of the lens can be demonstrated with the aid of the law of refraction and simple geometric reasoning.

If a point does not lie in the focus but at a distance a from the centre of the lens, then rays emanating from it collect at a distance a' These two distances are connected by the familiar formula

$$\frac{1}{a} \quad \frac{1}{a'} = \frac{1}{f}$$

where f is the focal length of the lens.

It is easy to show that light rays proceeding from an object twice the focal length produce an inverted and reduced (in the ratio of a'/a) image between the focus and the double focal length.

If the object is placed in the position occupied by the image, then the image takes up the position occupied by the object. This is the so-called *principle of reversibility* of light rays.

When we use a lens to magnify, the object lies between the lens and its focal point. In this case the image is not inverted and lies on the same side as the object.

The difference between the case of a magnifying glass and the two preceding instances is this: a magnifying glass produces virtual image, whereas in other positions of the object we obtain images that can be seen on a screen or photographed. We can justly call them real.

The magnifying power of a magnifying glass is the greater the smaller its focal length. The limiting possibilities of a magnifying glass are rather modest: the angle of view at which the virtual image is visible can only be magnified 20 to 30 times the angle of view at which we see the object with the naked eye.

Many optical instruments would be simple in the extreme and would consist of single lenses if it were not for a number of unavoidable defects. We want a parallel beam of white light to be focussed by a lens to a single point. But dispersion hampers this. Photons of different colour are deflected by the lens in different directions.

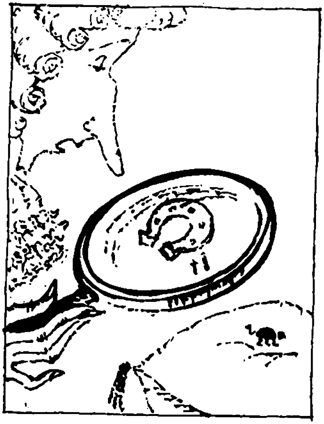


Figure 2.2

As a result, instead of a point we obtain a coloured line spread out along the axis of the lens. This is known as *chromatic aberration*.

Another bit of trouble is *spherical aberration*. The rays that are closer to the axis of the lens will come to a focus at a more distant point than rays whose paths lie farther away from the axis.

Also, the behaviour of rays falling on the surface of the lens at large and small angles is quite different. Instead of a point we obtain a glowing nucleus displaced away from the proper position. A tail-like appendage is attached to the nucleus. This effect is termed a *coma*. The word "coma" translated from the Greek is something in the nature of "loose hair".

And this is not the end of the list of distortions that plague the single lens. When we examine a square, we see a quadrangle with the vertices in the form of arcs convex inwards. This is because the rays emanating from the vertices of the square and from the midpoints of its sides are refracted in different ways.

Another defect of a lens that plagues designers of optical instruments is termed *astigmatism.* If a point lies a good distance from the principal optical axis of the lens, then its image splits into two strips perpendicular to each other and displaced in opposite directions with respect to the position of the ideal image.

And there are still other distortions. Specialists in lens production classify all types of distortions into seven basic types. We have mentioned only five.

As so often is the case in technology, making a good lens amounts to a compromise. It is quite clear that increasing the size of a lens increases the distortions, but, on the other hand, the illumination of the image (that is, the number of photons of visible light per unit area) is proportional to the square of the diameter of the lens (that is, its area). But this is not all. Suppose the object depicted by the lens is at a considerable distance. Then the image will come to a focus. The smaller the focal length the smaller the dimensions of the image. In other words, the light flux emanating from the object collects over a smaller area. Thus, the illumination is inversely proportional to the focal length.

For these two reasons, the square of the ratio of the diameter of a lens to its focal length is called the *aperture ratio* of the lens.

Thick lenses have the smallest focal lengths. These are lenses whose surfaces are formed by small radii. But these are precisely the lenses that produce the greatest distortions. This means that increasing the aperture ratio of a lens (whether done at the expense of its dimensions or the radius of curvature) leads to a worsening of the image. This is no easy task that confronts engineers and designers.

The Camera

In its simplest form, a *camera* is a lens that plays the role of a window in a dark box. The image produced by the lens is recorded on a photographic plate located opposite the window.

Now, a simple lens distorts any image. For this reason, it is replaced by a set of lenses designed to eliminate the optical defects encountered. The set of lenses is termed a *photographic lens*.

The question now is how to get around the distortions we have mentioned. Quite some time ago the idea was suggested to use a system of lenses chosen so that the defects of each one are eliminated by the defects of the others. This principle of obtaining a plus by multiplying two minuses has proved to be sufficient to eliminate all seven defects with the aid of only three lenses. That is the principle. Actually, to obtain the most perfect images requires still more complicated combinations. One such combination (though by far not the most complicated) is depicted in Figure 2.3. This system of concave and convex lenses is capable of producing a nondistorted image under appreciable variations of the degree of magnification. The first and third components of the system are capable of moving with respect to each other, thus attaining a continuous variable three-fold focal length.

A camera requires a rather simple device to "focus" it. This means making it possible to vary the distance between the centre of the lens and the film. Occasionally one sees old cameras in the form of an accordion that can be squeezed together. And I must say that such cameras do a fairly decent job.

In a modern camera that fits into the palm of your hand, this operation is performed more neatly: just a tiny spiral turn of the lens mount. As is evident from our discussion of the aperture ratio of a lens, the quality of

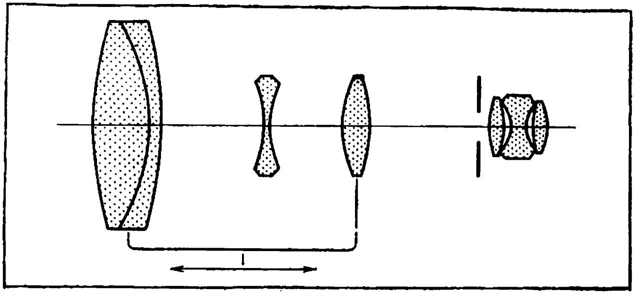


Figure 2.3

the image improves if we diminish the eye of the camera as much as possible. This is done with the aid of a diaphragm of variable diameter.

We choose the dimensions so that it is as small as possible but lets in sufficient light to produce a good image for the specified exposure.

Did you ever think of why the old photographs taken when photography was in its infancy look so stilted, so tense? The explanation is simple: the photographer had to resort to long exposures and that is why he exhorted his client not to move: "Hold it, ..."

The fight to get a good image at minimal exposure is carried out along two lines. The first is to perfect the photographic lens. This is done not only by suitable selection of the geometry of the lenses that make up the compound lens. In a compound lens made up of several lenses nearly half the light is reflected. First, this results in a loss of illumination of the image, second, it produces a light background that reduces the degree of contrast of the image. This is combatted by a technique called *lens coating*. The surface of each lens is covered with a very thin film. Due to the phenomenon of interference, the portion of reflected light is drastically reduced. Compound systems with coated lenses can be recognized at once: the glass has a bluish tinge.

Another way to improve photographs is to perfect the film.

A few words are in order concerning the photochemical process that leads to the formation of an image. The photosensitive layer is a gelatin with embedded crystals of silver bromide and a small admixture of silver iodide. The size of the crystal grains ranges from a thousandth to a ten-thousandth of a millimetre. The number of grains per square centimetre of film is anywhere from ten thousand to hundreds of thousands. Under the microscope, the layer of photographic emulsion reveals that the grains are rather close together.

Photons falling on a grain of emulsion disrupt the bonds between the atoms of silver and the atoms of the halide. The number of atoms of silver that are released is strictly proportional to the number of photons incident on the film. The photographer chooses an exposure time during which a considerable number of bonds between the atoms of silver and bromine are disrupted. And yet the exposure should not be too long. That would result in a complete destruction of the bonds between atoms of silver and bromine in all crystals. When the film is developed, the crystals release all the silver they contained and the film is black throughout.

If the exposure is correct, the photographic plate will reveal the latent image of the object. In each grain, the number of disrupted bonds is proportional to the number of photons incident on the grain. The process of developing the film consists in permitting the potentially free atoms of silver to combine. Then the amount of released silver on the negative after developing the film will be proportional to the intensity of the light.

From the foregoing it is probably evident to the reader that the smallest details revealed by a photograph of an object cannot be larger than the size of a grain crystal of silver bromide.

After the film is developed the next stage is to fix it. The fixing process consists in removing the undecomposed silver bromide. If these grains are not removed, then the film is spoiled when exposed to light because the grains release all the silver they contain.

The physics of obtaining a positive image is so obvious that we will not dwell on it.

The technology of modern coloured photography is not at all simple and merits our full admiration, yet the physics of that process, on the contrary, is very simple indeed. The model of our perception of colour that was proposed in the middle of the eighteenth century is quite true. The human eye has receptors of three colours: red, green, and blue. By combining these colours in different combinations, we obtain a perception of any colour. Accordingly, to obtain a coloured image we need a three-layer film. The upper layer is sensitive to blue rays, the middle layer to green, and the bottom layer to red. We will not go into how chemists achieve this. The coloured negative is transformed into a coloured positive, again through the use of three-layer photographic paper.

The Eye

The eye created by nature is marvelous physical instrument. The possibilities of distinguishing tens of thousands of shades of colour, of seeing close to and far away, of perceiving, via two eyes, the spatial relationships of objects, of being sensitive to extremely slight light intensities are all properties that place the human eye in the category of the highest-quality instrument. True, the human eye sees only a small portion of the spectrum.

The eyes of some animals, however, overcome this defect to a certain extent.

In some ways, the eye is reminescent of the ordinary camera. The role of the camera lens is played by the crystalline lens, which is double-convex. The crystalline lens of the eye is soft and capable of changing its shape under the action of muscles that embrace it. Therein lies the process of accommodation of the eye that permits it to see with equal ease both nearby and distant objects. With age, the crystalline lens becomes hard and the muscles weaken, and then glasses (spectacles) are needed for distance and reading.

The image of an object is projected onto the rear wall of the eye, and the optic nerve transmits this perception to the brain.

The normal eye of a young person is capable of discerning the details of an object located at a distance not less than 10 centimetres. Ordinarily, with age comes farsightedness, and that distance increases to 30 centimetres.

In front of the crystalline lens is the pupil, which plays the role of the diaphragm of the camera. The human pupil can vary in size from 1.8 to 10 millimetres.

The role of the photographic plate on which the image is formed is played in the human eye by the retina, which has a very complicated structure. Under the retina lies the optic epithelium, which consists of light-sensitive cells called rods and cones. You can compare the number of these cells to the number of grains of silver bromide in the photographic plate. There are over a hundred million optic cells in the human eye. Since the normal person is capable of distinguishing colours, it is clear that the optic cells possess differing sensitivity to different portions of the spectrum. We arrive at the same result if we assume that the cells are divided into classes receptive to different portions of the spectrum.

If vision is normal, the rear focus of the eye in the calm state lies on the retina. If it lies in front of the retina, the person is *nearsighted*; if it is behind the retina, the person is *farsighted*. These two most common defects of sight are due to an excessively thick or thin crystalline lens. Some people also suffer from *astigmatism*. In this case, the crystalline lens in the normal state does not have a correct shape bounded by two spherical surfaces.

All these defects can be rectified by eye glasses, which, together with the crystalline lens, yield an optical system that can focus the image of a distant object on the retina.

The lenses of spectacles are characterized by what is called *diopter* (a unit of the power of lenses). The optic power of a lens is inversely proportional to the focal length. The optic power in diopters is equal to unity divided by the focal length in metres. The focal lengths of dispersing lenses that nearsighted people use in their eye glasses are negative.

The angle of vision of the eye is much greater than appears to us. A number of events that occur at right angles (90°) to either side of our straightforward view are recorded directly by our subconscious. This fact leads some people to the erroneous idea that they "feel" someone is looking at them, without actually seeing the person.

The eye poorly perceives objects seen at an angle of less than one minute of arc, even if the illumination is very good.

Polarizers

A light wave is an electromagnetic wave. As was mentioned in the third book of this series, pictorial experiments have demonstrated that the vector of the electric field is perpendicular to the direction of motion of a ray of light. If this fact is interpreted from the standpoint of light being corpuscular (in the form of particles), then we must say that a particle of light—the photon—is not a sphere but an arrow. In a number of complicated calculations, theoretical physicists have come to the conclusion that the photon possesses spin (equal to 1). Thus, the concept of a photon as an arrow is quite natural.

The ordinary ray of light is a stream of photons whose spins are perpendicular to the direction of propagation of the light but are distributed uniformly in the form of a circle perpendicular to the light ray. This type of light ray is said to be *nonpolarized*. However, in a number of cases we have to do with a beam of photons in which all the spins are in one direction, or, to put it differently, we have to do with electromagnetic waves, the electric vectors of which have a very definite direction. Such rays are said to be *polarized*.

One way to obtain polarized rays consists in making the light ray pass through a low-symmetric crystal. Such crystals, when suitably oriented with respect to an incident light ray, possess the capability of splitting the natural ray into two rays polarized in two mutually perpendicular directions.

Unfortunately, I cannot give the reader even a rough idea about how this happens because the molecules of a crystal "receive" waves with differently arranged electric vectors in different ways. I can see that the preceding sentence hasn't helped matters much. But I can at least assure the reader that the theory of the splitting of light rays exists and it is a very good theory that is capable of describing all the fine details of this exciting phenomenon. For instance, it is possible to predict how the passage of light will change if we place the crystal at different angles to the ray of light.

By splitting a nonpolarized ray into two polarized rays, we can then easily make one of the rays go off in some desirable direction. The result will be what is called a *Nicol prism*, named after the Scottish physicist William

Nicol (1768-1851). The instrument was made in 1828. It is interesting to note that in those days all explanations of the polarization of light were given in the language of particles and it was considered an excellent confirmation of the corpuscular theory of light proposed by Newton.

Soon afterwards the interference and diffraction of light were discovered, and these were so naturally accounted for in terms of waves that the theory of particles of light was buried. But a century passed and the theory was resurrected, like the phoenix arisen from ashes. True, this time merely in the modest attire of one of two aspects of the electromagnetic field.

If a polarizer is placed in the path of the light, the intensity of the ray falls off, as is to be expected, by a factor of two. But the most interesting phenomenon, which is what proves the existence of polarization, occurs when we place a second instrument in the path of the light ray. This instrument is called an *analyzer*, although it does not at all differ from the first Nicol prism. Now let us turn the Nicol prism about the light ray. It turns out that for a certain mutual position of the two prisms, the intensity of the light that has passed through both Nicol prisms remains the same as it was in the absence of prisms. We then say that the Nicol prisms are parallel in that position. Now start turning the analyzer. When we have turned it through 90°, the light ceases to come through. We then say the Nicol prisms are crossed.

In an intermediate position when the second Nicol prism is turned from the parallel position through an angle α, the intensity is equal to $(1/2)\, I \cos^2 \alpha$. The formula is readily explainable if we assume that the vector of the electric field has been divided into two components, one perpendicular and the other parallel to the "slit" of the analyzer. Now the intensity is proportional to the square of the amplitude of the wave, that is, to the square of the electric vector. That is why variation in light intensity must occur in accord with the law of the square of the cosine.

Such an analysis of polarized light has a number of practical applications. Suppose the Nicol prisms are crossed and a transparent body that is capable of turning the electric vector of the wave is interposed between them. We then have field brightening. Charged bodies have this capability. Depending on the amount of voltage, the rotation of the light vector and, together with it, field brightening beyond the crossed Nicol prisms will be different. We will see pretty patterns, and coloured too, because photons of different colours behave differently. These patterns permit judging the voltages in the sample or deciding whether the molecules that make up the sample are oriented or not. This is important information, and so a good microscope is equipped with two Nicol prisms so that the image of an object can be seen in polarized light. The information about the structure is then much richer.

Solutions of many substances (for example, sugar solutions) have a certain rotatory power that permits turning the electric vector of a light wave. Here, the angle of rotation turns out to be strictly proportional to the amount of sugar in the solution. We can thus adapt a polarimeter for measuring sugar content. Such instruments are termed saccharimeters and can be found in any chemical laboratory.

These two examples do not exhaust the use of polarimeters but they are probably the main ones.

The Microscope and the Telescope

The optical part of a *microscope* consists of an eyepiece and an objective. The eyepiece (or ocular) is the lens to which the eye is applied; the objective is close to the object being studied. The object is placed at a distance somewhat greater than the focal length of the objective. The image obtained between the objective and the eyepiece is inverted. It is necessary that it appear between the eyepiece and the focus of the eyepiece. The eyepiece plays the role of a magnifying glass. It can be demonstrated that the magnifying power of a microscope is equal to the product of the magnifications of the eyepiece and objective taken separately.

At first glance it might appear that a microscope can be used to discern arbitrarily small details of an object. For example, why not make a photograph by increasing the dimensions thousands of times, then examine it in a microscope and obtain a million-fold magnification, and so on.

This kind of argument is totally fallacious. First of all, let us recall that increasing the size of photographic pictures is limited by the size of the grains of the film. The point is that each tiny crystal of silver bromide acts as a whole. The reader most likely has seen highly magnified photographs and has noticed that the magnification does not at all improve the details of the picture but only smears them out.

But if we discard the operation of photography and magnify the image optically, and this is possible merely by increasing the number of lenses, we will soon see that any great magnification is meaningless. The useful magnification of any instrument is limited by the wave aspect of the electromagnetic field. Whether we are examining an object through a magnifying glass with the naked eye or with the aid of a microscope or telescope, in all cases the light wave proceeding from the glowing point must pass through an opening. But in that case we have to deal with the phenomenon of *diffraction*, that is, the deviation of the light ray from its rectilinear path. To one degree or another, the ray "peeks around the corner" and so the image of the point is never a point but a spot. And no matter how you try, it is impossible to make the size of the spot less than the wavelength of the light.

It is essential to be able to estimate under what conditions the path of the electromagnetic wave departs appreciably from the rectilinear path.

If x is the linear deviation from the rectilinear path observed at a distance f from the source of radiation, and the size of the obstacle (or aperture) in the path of the ray is equal to a, then we have the following relationship:

$$x = \frac{\lambda f}{a}$$

Here, λ is the wavelength. From this equation it follows that diffraction can be observed both in the case of minute particles and celestial bodies. The whole point is the wavelength and the distances we are dealing with. The same goes for apertures as well. Diffraction is not only observed in the case of minute openings. For example, an opening the size of a tennis ball permits observing diffraction phenomena; true, only at distances of the order of hundreds of metres.

The simple equation we wrote down enables us to gauge the limiting possibilities of microscopes and telescopes.

A microscope does not permit us to see features of an object smaller than a micrometre. Now millimetre sizes can be seen with the naked eye. This means that optical microscopes should not be used to seek magnifications greater than 1000.

This restriction has to do with optical microscopes only. Now if it were possible to design a microscope operating not with light rays but with some other rays with smaller wavelengths, then the useful magnifying power of the microscope could be increased. Just such a microscope exists and can be found in many scientific labora-

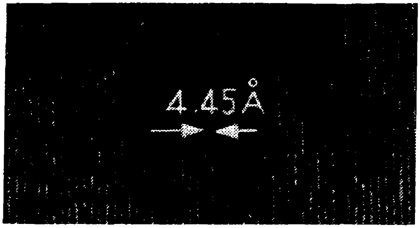


Figure 2.4

tories. It is called an *electron microscope*. The wavelengths of electrons can be made very small (see page 129). Electron microscopes can resolve structural details of the order of ten millionths of a millimetre.

Biologists have seen molecules of DNA, the long molecules that carry hereditary characteristics from parents to progeny. Protein molecules have been observed, also the structure of cell membranes and details in the structure of muscle fibres. The photograph in Figure 2.4 is a record: this is a 3-million-fold magnification of the crystal lattice of the mineral pyrophyllite showing the distance between the planes of the crystal at 4.45 angstroms!

The limit of the electron microscope has to do not with its resolving power (we can easily reduce the wavelengths of electrons) but with the contrast in the image: the molecule under investigation has to be backed: placed on a base (backing), which itself consists of molecules. On the background of molecules of the backing, it is difficult to discern the molecule of interest.

An electron microscope is a complex and very expensive instrument about a metre and a half in height. Electrons are accelerated by a high voltage. The principle of magnification is the same as in the optical microscope. It is produced by lenses. But, naturally, they are not at all like those of an ordinary microscope. The electrons are focussed with the aid of electric fields of high-voltage metallic plates with apertures and by means of coils that set up a magnetic field.

There are a great variety of techniques that help to produce the image. Microtomes are used to obtain extremely thin sections for transillumination, the molecules on the backing are tinged by depositing metallic vapour on them. It is also possible to obtain a replica of the specimen, that is, by coating it with an extremely thin film of transparent material and then etching out the object itself.

Electron microscopy is a big and important division of physics well deserving of a separate chapter, but the size of this book presses me onwards.

Our next topic is the *telescope*. As early as the 16th century suggestions were made that distant objects could be examined with the aid of convex glasses. Yet I think that the real invention of the telescope (actually a spyglass) was made by the great Galileo. He constructed it in July 1609 and a year later published his first observations of the night sky.

Like the microscope, a telescope (*refracting telescope*, to be exact) is in principle a combination of two lenses: the objective, that forms an image from a distant object, and the eyepiece. Since an infinitely distant object is viewed, its image is created in the focal plane of the objective. The focal plane of the eyepiece coincides with the plane of the objective, and beams of parallel rays emerge from the eyepiece.

The power of the telescope increases with the diameter of the objective. For example, a 10-cm-diameter telescope permits viewing features on Mars about 5 km in size. With a 5-metre diameter telescope, objects up to 100 metres across can be resolved on Mars.

Astronomical observatories are equipped not only with refracting telescopes but also *reflecting telescopes*. Since our aim is the study of distant objects and it is required to gather the light rays into a focus, this can be done by means of a spherical mirror as well as a spherical lens. The obvious advantage in this case (a mirror) is that we get rid of the chromatic aberration. The defects of the mirror telescope are connected only with the difficulty of making high-quality mirrors.

Quite naturally, the telescope too has a limit to the useful magnifying power it is capable of producing. This limit is associated with the wave nature of light. A ray of light from a distant star spreads out into a circle and this places a limit on the angular distance between stars that we are able to resolve in the telescope. Again, the desire to increase the power of the telescope is linked up with increasing the diameter of the glass. The limiting possibilities of the telescope probably lie somewhere close to one tenth of a second of arc.

During recent years, new instruments have appeared and are helping astronomers study the stars in all portions of the spectrum of electromagnetic waves coming to us from outer space. We will come back to this topic again in the seventh chapter.

Interferometers

As we have already pointed out a number of times, the electromagnetic field possesses a wave aspect. And so also do beams of particles—electrons, neutrons, protons. Sound is the result of mechanical displacements of the medium and these occur in accord with the wave law. Common to all these physical processes is the possibility of ascribing to any radiation a wavelength, frequency, and rate of propagation; these are connected by the equation $c = \lambda\nu$. The simplest kind of radiation is monochromatic, that is, it is described by a single wavelength. In the general case, radiation constitutes a complex spectrum, that is, a sum of waves of different wavelength and intensity.

The wave aspect of radiations is revealed in the combining of waves and also in scattering due to bodies in the path of the rays. An important special case of wave scattering is diffraction. The composition of waves is *interference*.

Here we will deal with the interference of light. This phenomenon underlies the operation of a number of instruments used for measuring distances and also certain other physical quantities. Instruments that use the phenomenon of interference for applied purposes are called *interferometers*.

The principle for measuring distance consists in calculating the number of waves that fit into the distance being measured.

This might appear to be a simple thing to do. Take two sources of light and bring both beams to a single point. Then depending on whether the waves arrive at the point of observation "hump to hump" or "hump to dip", we get a bright or a dark spot. Now let us pose the problem of measuring the distance over which we wish to move one of the light sources. This motion will cause the phase relationships of the two waves at the point of observation to change. All we have to do is calculate the number of alternations of light and darkness and then, taking into account the geometry of the experiment and knowing the wavelength of the light, readily compute the amount of displacement.

In principle that is correct. But acting in this fashion we will not observe a pattern of alternating light and darkness. The screen will remain light all the time and so this simple experiment is a failure.

What is definitely known is that if two rays of light emitted by two different sources come to a point, they will always reinforce each other. Then perhaps the wave theory is fallacious?

No, the theory is correct and electromagnetic radiation has a wave aspect. But we began with an erroneous supposition. For interference to be observed, it is necessary that an invariable phase difference obtain all the time between the waves being combined. But phase relationships even between waves emanating from two atoms of one and the same source are quite accidental. We have already mentioned the fact that atoms throw out photons without "agreeing" among themselves on their behaviour. Thus, two different sources radiate haphazardly; this is called *incoherent radiation.*

But then coherent radiation must be something in the nature of a dream. We shall now see that that is not so.

The solution is extremely beautiful and at the same time simplicity itself, like most original concepts: the problem is to make the radiation of the atom combine with itself. And to do that it is required to split the ray coming from each source into two parts and make the two parts of the single ray traverse different paths, and finally bring them together. Then, under such conditions while observing interference and changing the differences in the paths of the parts of the split ray, we can indeed measure the displacement and length by computing the number of alternations of light and darkness.

What we have just described is the principle underlying interferometer measurements that was discovered in 1815 by the French physicist Augustin Jean Fresnel (1788-1827). Let us now consider the methods underlying the action of interferometers by means of which one can split a ray of light and create different paths between the split portions of the ray.

Now let us dwell in more detail on the interference of light rays reflected from the inner and outer sides of a transparent plate or film. This is of interest both because of its practical utility and because it is observed in nature. Besides, this is a good example for illustrating many

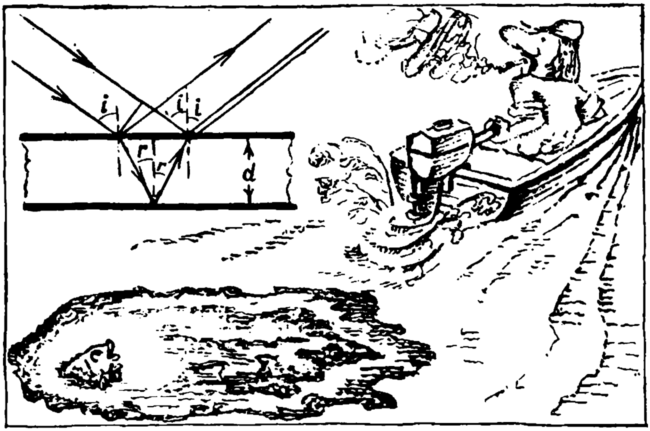


Figure 2.5

important concepts that we use in describing light waves and other electromagnetic waves.

Figure 2.5 enables us to compute the phase shift between two such rays. The difference in phase is determined by the difference in routes, that is, the difference in the paths taken by the two rays. From the drawing it is evident that the path difference is $x=2d \cos r$. But then there is the question of how to proceed from the path difference of the rays to the phase difference, which is what determines whether the two waves will reinforce or weaken each other.

If the reader is not afraid of the cosine formula, we can discuss this matter. The oscillation of a light vector at any point in space can be written down as follows: $A \cos 2\pi\nu t$. The phase shift through the angle φ signifies the necessity of adding this angle to the argument of the cosine. If we want to compare the phases of points of one and the same wave, which points are separated by a distance x, we have to take into account the number of wavelengths that fit into that interval and then multiply the result by 2π. That quantity is the *phase shift*. Thus, $\varphi = 2\pi x/\lambda$.

Now let us return to the interference of rays in the plate. We have written down the difference in paths. All that is left, then, is to divide that quantity by λ. But wait a minute. Is the wavelength of light in a vacuum the same as inside a transparent plate? Something makes us suspect that when a wave of light passes from one medium into another it undergoes a change. We know about *dispersion*—when photons of different frequency behave differently. The frequency, wavelength, and rate of propagation are related as follows: $c = \nu\lambda$. Which of these quantities undergoes a change when the wave enters a new medium? Experiment can tell us. We can measure the rate of propagation of the wave directly in the body and then see that the refractive index, which makes the wave alter its direction of motion when falling obliquely on the interface between the two media, is equal to the ratio of the rates of propagation of light in the media. If one of the media is air (a vacuum, to be more exact), then

$$n = \frac{c}{v}$$

where c is the velocity of light in a vacuum and v is the rate of propagation in a medium. Now the question is which of the two parameters (frequency or wavelength) undergoes a change when light moves from the air into the medium. To account for the results of the interference experiments, we have to assume that the frequency of the photon remains unchanged and the wavelength changes. That is why the following formula holds true for the refractive index:

$$n = \frac{\lambda_0}{\lambda}$$

where λ_0 is the wavelength in air.

Now we know all we need to in order to write down the phase difference between the rays in the plate experiment. Since one of the light rays moved in the air and the other in glass, the phase difference is

$$\varphi = \frac{2\pi}{\lambda_0} nx = \frac{4\pi}{\lambda_0} nd \cos r$$

What can we measure by studying the interference of light rays in a plate? The formula answers that question. If we know the thickness, we can determine the refractive index of the material. If we know n, we can find the thickness to a high degree of accuracy (fractions of a wavelength of the light) and, finally, we can measure the wavelengths of different colours.

If the plate is of variable thickness and if the material is homogeneous, and the angle of incidence is practically the same for the portion of the plate we are interested in, then the interference will be exhibited as so-called bands of equal thickness. On an uneven plate, we have a system of dark and light bands (they may be all the colours of the rainbow in the case of white light because the photons of each colour will behave in their own distinctive way), the bands revealing places of equal thickness. That is the explanation of the varicoloured streaks that we see on films of oil spilt onto water.

Soap films exhibit very pretty bands of equal thickness. Make a simple wire frame, dip it into sudsy water, and pull it out. The soap suds slide off and the film in the upper part is thinner than it is in the lower portion. The film exhibits coloured horizontal bands.

Interference methods are widely employed in measuring small distances or small changes of distance. They permit measuring changes in thickness less than hundredths of the wavelength of the light. In interference measurements of uneven places on the surface of a crystal, precision of the order of 10^{-7} centimetre has been attained.

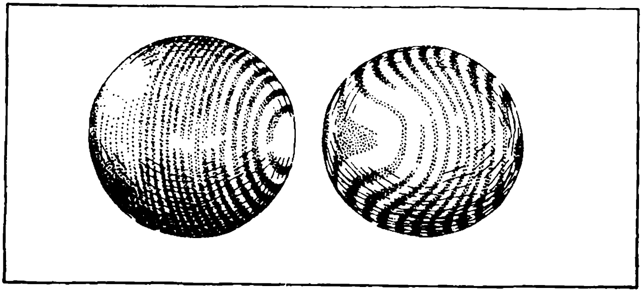


Figure 2.6

This method is widely used in optical work. For example, if we need to check the quality of the surface of a glass plate, this is done by considering the bands of equal thickness of a wedge of air set up by the plate with a perfectly smooth surface. If the two plates are pressed together at one end, a wedge of air is produced. If both surfaces are plane, the lines of equal thickness will be parallel lines.

Suppose the plate being tested has a depression or a bump. The lines of equal thickness will be bent because they bend around any defective spot. As the angle of incidence of light is changed, the bands move to one side or the other depending on whether the defect is a bump or a depression. Figure 2.6 illustrates what the microscopic field looks like in such cases. Both patterns reveal defects in the samples. In the first case, the defect is located on the right edge, and in the other, it is on the left edge.

Exact measurements of refractive indices of a substance can be made with the aid of interference refractometers. These instruments reveal interference between two rays that have the largest possible separation.

Suppose in the path of one of the rays we have a body of length l and refractive index n. If the refractive index of the medium is n_0, the optical difference of the path changes by $\Delta = l\,(n - n_0)$. The two rays are brought to a single point by means of a focussing lens. What will we see in telescope? A system of light and dark bands. But these are not bands of equal thickness that can be seen with the unaided eye. The system of bands appearing in a refractometer is of a different origin. This is because the original beam of light is not ideally parallel but slightly divergent which means the light rays constituting the cone will fall on the plane at slightly different angles.

The interference events will occur in the same manner in the case of rays having the same inclination. And they will come together in the same spot of the focal plane of the telescope. If the path difference between the split portions of the beam changes, the bands will begin to move. If the path difference changes by the amount Δ, then Δ/λ bands will pass through the eyepiece of the telescope.

The method is extremely precise because a displacement of 0.1 band can be detected with ease. For such a displacement, $\Delta = 0.1\lambda = 0.5 \times 10^{-5}$ centimetre, and this permits detecting, over a length $l = 10$ centimetres, a change in the refractive index amounting to 0.5×10^{-6}.

We will now examine an interferometer of a different kind that does not utilize the phenomenon of refraction. This interferometer was made by the American physicist Albert Abraham Michelson (1852-1931). It is hard to overestimate the role this instrument played in the history of physics (I make bold to say, even the history of human thought). Through the use of this interferometer, a fact of exceptional importance was discovered: the speed of light in directions along the earth's orbit and across it is the same.

What this means is that the speed of light does not depend on the motion of the light source! This means

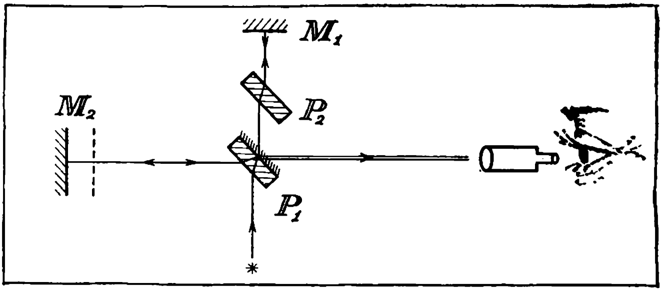


Figure 2.7

that the speed of light is not combined with the speed of motion of the lamp (that produces the light flash) in accord with the rules by which the speed of a bullet is combined with the speed of the man holding the gun. The discovery of this remarkable fact led to the theory of relativity and to a fundamental reconsideration of the meaning of such basic scientific concepts as length, time, mass, and energy. We will go into these important matters later on. The *Michelson interferometer* is of interest not only because of the place it occupies in the history of physics but also because the simple principles underlying its construction are still used today to measure lengths and distances.

In this instrument, a parallel beam of monochromatic light falls on a plane-parallel plate P_1 (Figure 2.7) covered with a translucent layer of silver on the ruled side. This plate is placed at an angle of 45° to the light ray falling from the source and splits it in two, one ray moving parallel to the incident ray (to mirror M_1) and the other perpendicular to mirror M_2. The separated rays of light fall on the two mirrors at zero angles of incidence and return to the very same spots of the translucent plate from which they emerged. Each ray returning from the mirror is again split on the plate. Part of the light returns to the source and the other part enters the telescope. In the figure you can see that the ray coming from the mirror opposite the telescope passes twice through the glass plate with its translucent layer. For this reason, in order to ensure equality of the optical paths, the ray coming from mirror M_1 is passed through a compensating plate P_2 that is identical to the first plate but without the translucent layer.

In the field of view of the telescope, circular rings are observed that correspond to interference in the layer of air (the thickness of which is equal to the difference between the distances of the mirrors from the site of beam-splitting) of the primary rays forming the cone. A displacement of one of the mirrors (for instance, M_2 to the position shown by the dashed line) by a fourth of a wavelength will correspond to a transition from maximum to minimum, that is, it will cause a shift in the pattern through half a circle. This can be clearly detected by an observer. Thus, in violet rays the sensitivity of an interferometer is better than 1000 Å (angstroms).

The advent of lasers produced a revolution in the technique of interferometry.

Did you ever wonder why an interference rainbow appears on films of oil floating on water but does not appear in window glass? The situation would seem to be quite similar: in both cases, rays are reflected from an outer and an inner surface of a plate. In both cases, the two reflected rays combine. But interference is observed only when the plate is thin. I've used this question at examinations and confused many a student.

The essence of the matter is this. The emission time of an atom is equal to 10^{-8} or 10^{-9} second. The single act of emission consists in the release of a wave train. Since the time of emission is very small, the wave train is very short despite the great velocity of light. When we split a ray into parts, only two parts of the same wave train can undergo interference. This means that one portion of the sine curve must appreciably overlap another portion. But for this to happen, it is naturally necessary that the path difference between the split portions of the ray be substantially less than the length of the wave train. The condition is met in the case of a micron-thin film, but not in the case of window glass. That is the answer the student should have given.

The maximum path difference between the rays when interference can be observed is termed the *coherence length.* For light, this amounts to a fraction of a millimetre.

Now notice how radically the situation changes in the case of laser emission. A continuous-action laser creates photons of stimulated emission that set out in the same phase. Or, to use wave language, the wave trains emanating from different atoms are superimposed on one another thus creating what appears to be a single wave. The coherence length becomes practically unbounded; at any rate, it is measured in metres and kilometres (as is always the case, the ideal is never attainable but I will not dwell here on the variety of factors that affect the coherence length).

Using laser light, it is possible to construct interferometers capable of handling problems that until recently had appeared unsolvable. For example, using an ordinary light source, the mirror in the Michelson interferometer can be displaced only distances of the order of a millimetre. Now if the light ray is produced by a laser, the path of the ray falling on mirror M_1 may be made equal to several centimetres and the path of the ray reflected from mirror M_2, tens of metres.

Interferometers to check the sphericity of lenses can be made with a single surface of comparison whereas, if we use ordinary light, one has to change the standard of comparison along with any change in the radius of the lens being tested (this is because one cannot work with large path differences). Which is to say nothing about the fact that the interference patterns have become incomparably brighter and for that reason can be analyzed more easily and exactly.

The possibility of getting along without compensation of the optical path of one of the rays makes it possible to design interferometers of an entirely new type. It has now become possible to detect displacements of dams, geological drift, oscillations of the earth's crust. By reflecting laser light from objects located at large distances and then making it interfere with the original light, we can carry out exact measurements of the speed of motion of such objects.

Laser Devices

A device that produces a laser beam can of course be called an instrument because it is used for analysis, control, and observations. However, unlike other optical instruments, the laser plays a far greater role in industry. The use of lasers is so widespread that we will again and again come back to them. In this section we will discuss the use of lasers in working various materials. If it is not high power that is required, use can be made of the compact neodymium laser. The heart of this laser is, as we have already mentioned, glass alloyed with neodymium. A glass rod 50 millimetres long and 4 millimetres in diameter. The light flash that does the pumping is given by a xenon lamp. In order to cut losses of luminous energy, the lamp and rod are enclosed in a water-cooled cylindrical chamber.

The following properties are important in the use of this type of instrument and its analogues: the possibility of localizing energy over an extremely small area, the possibility of precise batching of portions of energy, and the possibility of delivering energy without the use of any wires or contacts.

Take the use of lasers in the watch industry. A watch requires jewels. These are often in the form of rubies; the more rubies the higher the quality of the watch. Apertures have to be drilled in tiny ruby discs. Ordinarily, without lasers, this operation takes several minutes for each jewel. Using lasers, the process is completely automatic and takes a fraction of a second. And since industry uses many millions of these a year, the obvious gain in time is tremendous.

The laser is extremely important also in the diamond industry. Drilling and broaching of diamonds is easily handled by lasers which can make a diamond into any shape or produce apertures as small as a few micrometres.

But let's get back to watches. A laser can weld the spring to the watch mechanism. Quite obviously, wherever high-precision welding is required—and this is the order of the day in many areas of high-precision technology—laser beams are employed with excellent results. The great advantage of this needle-thin ray is that there is no need to protect and cool adjacent parts of the component being welded.

Lasers are of course used trivially as a knife to cut out any imaginable patterns on any kind of material.

Lasers have invaded some rare professions as well. Take the restoration of marble sculptures. The atmosphere of the twentieth century is unfortunately quite polluted. A great variety of noxious gases, particularly sulphur oxide, form a black crust on marble. This crust is porous and acts as a sponge packing up more moisture and more harmful substances. Removing the crust by mechanical or chemical means can often ruin the sculpture. Now a laser operating in a pulsed regime can remove the crust without harming the underlying marble.

A carbon dioxide laser can be used to grow crystals without resorting to the crucible. This is not a new procedure. High-frequency currents have long been used in this capacity, but not for dielectric materials that possess too low a thermal conductivity. Today, lasers are used to grow crystals (without using crucibles) of niobates and other much needed materials. The importance of the noncrucible growth of crystals for the needs of microelectronics cannot be overestimated because even a few parts in a million of impurities can play a negative role, and in a crucible it is practically impossible to keep harmful atoms from passing from the material of the crucible into the crystal.

I will not describe the apparatus used in such cases. Crystal growth was discussed in the second book of our series. As in the case of high-frequency currents, the laser beam creates a small molten zone that slowly brings the material up to the growing crystal. I think it very probable that lasers will be supplanting other methods of crystal growth as well.

Photometry

Every source of light may be characterized by the energy it emits. However, in many cases we are only interested in that part of the energy flux that produces a visual sensation. As we have already mentioned, this is characteristic of electromagnetic waves of wavelength between roughly 380 and 780 nanometres.

Light perceived by our brain is characterized by luminance (brightness) and colour. If we compare the visual sensations created by light of equal intensity but different wavelength, it will be apparent that the eye perceives as brightest light that has a wavelength of 555 nanometres. This is green.

The perception of light is shown as a *visibility* (or *luminosity*) *curve* in Figure 2.8, which represents (in

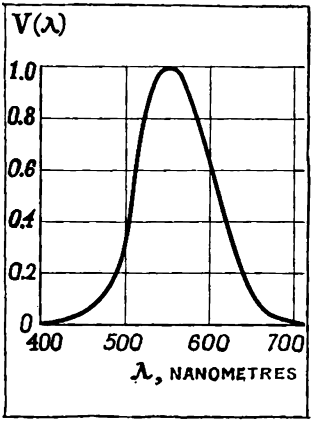


Figure 2.8

relative units) the sensitivity of the normal human eye to waves of different length. However, engineers disregard this curve and leave it up to the human eye to judge the integral *luminous intensity*. The first thing, in that case, is to choose a standard light source and then compare other sources with the standard. For a long time, the unit of luminous intensity was the *candle* because at first attempts were made to find a certain standard candle flame. It is quite obvious that this is no easy job.

The international standard today is an incandescent black body. The material is platinum. The black body emits light radiated through a small aperture by platinum heated to the melting point, or 1769 °C.

The unit of luminous intensity is the *candela* (which is Latin for "candle"). The international definition avoids any direct indication as to temperature of luminescence (this is done to avoid errors associated with measuring the temperature). The candela is defined thus: if we take, as the source, platinum as it solidifies at normal atmospheric pressure, then an area of $(1/6)\times 10^{-5}$ square metre yields a luminous intensity, in the direction perpendicular to the surface, equal to one candela.

The definition of luminous intensity and its unit are surprisingly archaic. Specialists realize this and they hope that in the near future a new basis will be chosen: energy measurements and the standard luminosity curve. Then luminous intensity can be defined by multiplying these two spectral distributions. If we multiply these two functions and integrate the product, we obtain a measure of luminous intensity. For the reader who is not acquainted with integral calculus I can restate this as follows: multiply the intensities of the separate spectral portions by their luminosity factors and then combine all the products.

At sufficiently large distances, a light source appears as a point. That is just when it is convenient to measure luminous intensity. Now, around this point source let us construct a sphere and on the surface of the sphere a portion of area S. Dividing S by the square of the distance from the centre, we obtain the so-called solid angle. The unit of the solid angle is the steradian. If we cut out an area $S=1$ square metre on a sphere of radius one metre, the solid angle is equal to one steradian.

Luminous flux is the luminous intensity of a point source multiplied by the solid angle.

Do not be upset by the fact that the luminous flux becomes zero (vanishes) when we speak of parallel rays. That is because the concept of luminous flux is not used in such cases.

The unit of luminous flux is the *lumen*, which is equal to the flux delivered by a point source with a luminous intensity of one candela to an angle equal to one steradian. The overall (total) luminous flux emitted by a point in all directions is equal to 4π lumens.

Luminous intensity characterizes a light source irrespective of its surface. Yet it is quite obvious that the impression will differ depending on the extent of the source.

For this reason, use is made of the notion of *luminance* (*brightness*) of a source. This is the luminous intensity per unit surface of the light source. Luminance is measured in *stilbs*: one stilb is equal to one candela per square centimetre.

One and the same light source brings different luminous energy to the page of an open book depending on where the source is located. What is important to the reader is the illumination of the portion of his desk in which his book lies. *Illumination* is defined as the luminous intensity divided by the square of the distance from the point source. Why the square? Because a luminous flux remains unchanged inside a given solid angle, no matter how far we recede from the glowing point. Now, the area of the sphere and the area of the portion cut out by the solid angle will increase in inverse proportion to the square of the distance. This simple rule is called the *inverse-square law*. If the distance of the book from a small lamp is increased from 1 metre to 10 metres, we thus reduce the illumination of the page by a factor of one hundred.

The *lux* is the unit of illumination. This is the illumination generated by a luminous flux equal to one lumen over an area of one square metre.

The illumination on a moonless night is equal to 0.0003 lux. On a night with full moon the illumination can reach 0.2 lux. For comfortable reading we need an illumination of 30 lux. When shooting movie scenes, powerful projectors are used that raise the illumination to 10 000 lux.

So far we have not mentioned the instruments that measure luminous fluxes and illumination. At the present time, such measurements are no problem. Actually, we proceed exactly as we did in providing the new definition of the candela. We measure the energy falling on a photocell and graduate the scale of the photocell in units of lux with account taken of the luminosity curve.

Photometers in use last century operated on the principle of comparing the luminances (brightnesses) of two adjoining illuminated areas. One area received light that was to be measured; on the other, using simple devices, one reduced the luminous flux a known number of times until the two adjoining areas had the same illumination.

Holography

The invention of lasers has opened up a new chapter in the development of science and technology. It is hard to find a branch of knowledge in which stimulated emission has not revealed fresh vistas.

In 1947, Dennis Gabor (1900-) outlined the principles of obtaining images of objects in a radically different way from photography. The new technique is called *holography*. Holography became possible because of the peculiarities of stimulated emission that make it different from ordinary light. Let us stress once again that in laser radiation nearly all photons coincide in all their features: frequency, phase, polarization, and direction of propagation. A laser beam is hardly at all spread out, which means an extremely narrow beam can be thrown to great distances from the source; laser beams have the property of a very great coherence length. It is due to this circumstance (which is particularly important in holography) that interference of split rays with a large path difference is possible.

The upper part of Figure 2.9 illustrates the technique of obtaining a hologram. The object of interest is illuminated with a broad and weak (so as not to harm the object) laser beam. The same beam is scattered by the object and by a mirror, which generates a so-called reference wave. The two waves are superimposed, interference occurs, and the pattern is recorded on a photographic plate.

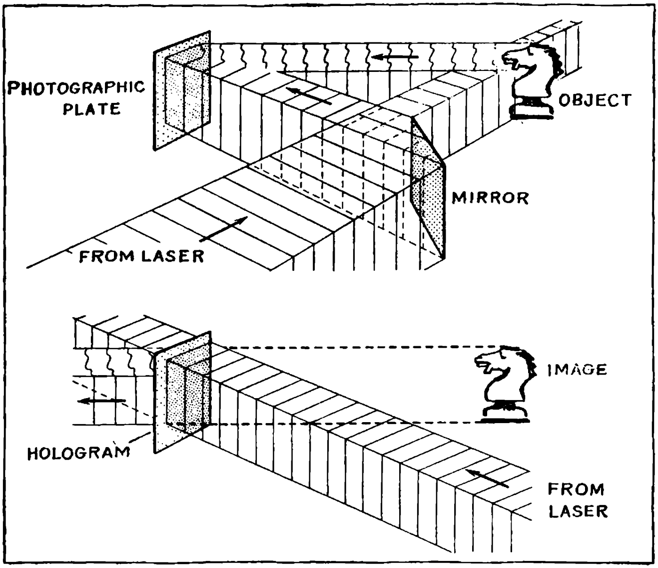


Figure 2.9

Take a look at Figure 2.10. The object is shown above, the "image" below. Yes, this is the image. It is an intricate combination of dark and light rings called a *hologram* and is indeed the image of the object, only it is a latent (hidden) image. The hologram contains complete information about the object; it would be more correct to say complete information about the electromagnetic wave scattered by the chess pieces. A photograph does not contain so much information. The best photograph conveys precisely all information about the intensity of the scattered rays. But a wave scattered by any point of an object is fully characterized not only by its intensity (amplitude) but also by its phase. A hologram is an interference pattern and each light or dark line describes

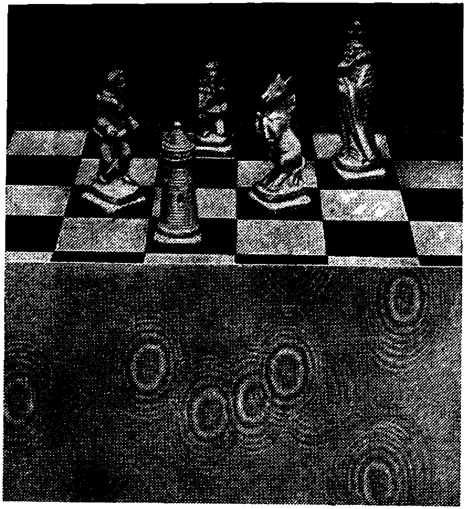


Figure 2.10

not only the intensity but also the phase of the rays coming from the object onto the appropriate spots of the photographic plate.

A hologram, like any ordinary photographic plate, can be developed, fixed, and stored for any length of time. Whenever we want to, we can view the object by irradiating the hologram, as shown in the lower part of Figure 2.9, with light from the same laser, thus reconstructing the original geometric arrangement: the laser beam is directed like the beam reflected from mirror. Then an image of the object appears where the object once stood, and that image, ideally, is identical to the picture that confronted the eye.

We will not go into the theory of obtaining holograms. The basic idea is that when a hologram is illuminated, there arise scattered waves with the same amplitudes and phases as those that created the hologram originally. These waves combine to form a wave front that is identical to the wave front that produced the hologram. What occurs is a peculiar reconstruction of the wave when the hologram is illuminated in the same conditions that the object was illuminated. The result is an image of the object.

Research in the field of holography continues. We are now able to obtain coloured images. Improved results are possible by taking several holograms from different angles. Finally (and this is perhaps the most important thing) it turns out that we can examine holograms without resorting to lasers.

There are books which treat the subject of holography in detail. Holography merits our attention for the reason that it is high-capacity method for storing three-dimensional information concerning an object. Future research will undoubtedly bring holography into new fields of technology and into the home.

3. Hard Electromagnetic Radiation

The Discovery of X Rays

X rays is the term used for radiation occupying a portion of the electromagnetic spectrum approximately from several tens of nanometres to 0.01 nanometre. The still harder (that is, having still shorter wavelengths) rays are called *gamma rays*.

As we have already mentioned, the names for the various portions of the electromagnetic spectrum are rather conventional. The term designating each portion of the spectrum has less to do with the wavelength than with the nature of the source of radiation. Most frequently, the term "X rays" is used for radiation that occurs when a flux of electrons encounters an obstacle.

Wilhelm Conrad Roentgen (1845-1923) discovered this kind of radiation on November 8, 1895. During those years, many physicists around the world were studying fluxes of electrons arising in evacuated glass tubes (illustrations of these are given in Figure 2.6 of the third book of this series). Two electrodes were soldered into a vessel and a high voltage applied. The fact that some kind of rays were emitted from the cathode of such a tube had been suspected for quite some time. At the very beginning of the 19th century a number of researchers observed flashes inside the tube, a fluorescence of the glass. Through the experiments of the German physicist Johann Wilhelm Hittorf (1824-1914) and the English chemist and physicist William Crookes (1832-1919) it was demonstrated that these were rays, in the path of which an obstacle had been placed. All textbooks included a photograph of the *Crookes tube* with a cross which he made in 1878, nine years after Hittorf. The cross threw a definite shadow onto the glass. This elegant experiment proves that some kind of rays emanate from the cathode and move in straight lines. When the rays fall on the glass, it glows and a thin layer of metal absorbs the radiation.

The fact that cathode rays are a flux of electrons was proved by Sir Joseph John Thomson (1856-1940) in 1897. By a method discussed in the third book of this series he was able to determine the ratio of the charge to the mass of the electron. Another 10 to 15 years passed and it became clear that the electron is a minute particle of electricity.

But we are going astray. What interests us now is the discovery made by Roentgen. However, our brief digression was made to emphasize the fact that Roentgen's discovery preceded an understanding of the nature of the rays emitted by the cathode. Actually, it was precisely because of a lack of understanding that Roentgen was investigating a variety of tubes with different locations of electrodes and different forms of glass envelope.

The events of the evening of November 8, 1895 are perfectly well known. Roentgen placed a piece of black cloth over the tube, turned off the light in the room, and was about to go home with the light switch left on. Then he glanced at the instrument he was working with and noticed that the screen covered with barium platinocyanide (which is capable of fluorescing) next to the tube was glowing. Roentgen returned, threw the switch, and the fluorescence ceased. He turned on the switch and the screen began to glow anew. Roentgen knew that in the tube of the design he was working with, cathode rays could not pass through the protective covering thrown

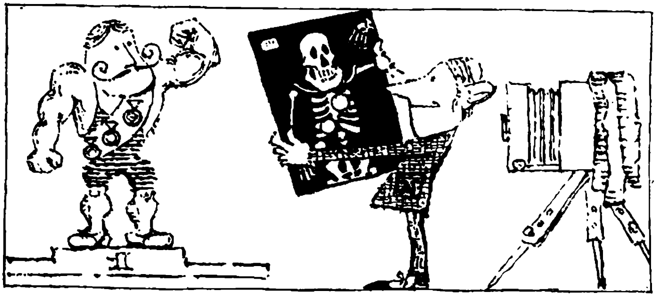


Figure 3.1

over the tube and then through a big layer of air as well. That could only mean one thing—a new hitherto unknown kind of radiation.

Roentgen first reported his discovery at the end of the year. During this interval he had made such a thorough study of the properties of the new rays that up until the discovery of the diffraction of X rays by a crystal (1912)—we will come to that soon—nothing new about the X rays was unearthed. The name X rays was given by Roentgen himself, but in some countries the radiation is termed Roentgen rays.

The most remarkable property of X rays is the property that Roentgen studied and illustrated from the very start: their ability to pass through materials that are opaque to light. (Figure 3.1 is a reminder of the many caricatures that appeared two or three months after Roentgen's first publication.)

The penetrating power of X rays was a boon to medicine. It also became possible to detect imperfections in industrial goods. The amazing results of roentgenography stem from the fact that materials of different density absorb X rays to different degrees. The lighter the atoms of the material the less the rays are absorbed.

It was also soon found that the penetrating power of the rays increased with increasing voltage in the tube. Voltages ordinarily used in X-ray equipment are between several tens and several hundreds of kilovolts.

Research into the properties of X rays elicited the fact that they are caused by retardation of electron fluxes due to an obstacle. It is curious thing to recall that for quite a long time Roentgen tubes were manufactured with three electrodes. An "anticathode" was fitted opposite the cathode to receive the impinging electrons. The anode was placed to one side. A few years later it became clear that that was a needless complication and today two electrodes are used. A flux of electrons impinges on the anode, the surface of which is beveled. Then the flux of X rays is directed where needed. If the surface of the anode meets the electron flux at right angles, the rays go off the anode in all directions and we lose a certain amount of intensity.

The use of X rays created a revolution in industry and medicine. X-ray technology today is a highly refined field. By varying the position of an object relative to the X-ray tube, we can obtain several pictures, thus enabling us to establish the exact position of a defect not only in projection but also its depth within the material.

Depending on the materials or tissues being investigated, use is made of hard (more penetrating) radiation or soft radiation. The main purpose is to attain a contrast so as to see the defect, which may differ in density only slightly from the main material.

The *law of absorption of X rays*, like the law of absorption of any other radiation, is quite obvious. What interests us is how the intensity of radiation (recall that intensity is energy per unit time and unit area) changes after passing through a plate of thickness d. Since this book is for the reader who is not acquainted with integral calculus, I will confine myself to the statement of the law for radiation passing through thin plates. "Thin" stands for the intensity falling off very slightly, say about 1%. In that case the law is simple: the portion of absorbed radiation is directly proportional to the thickness of the plate. If the intensity diminishes from a value I_0 to a value I, then this simple rule can be written thus:

$$\frac{I - I_0}{I} = \mu d$$

The proportionality factor μ is called the *absorption coefficient*.

Here is a simple question I have posed at examinations: In what units is the absorption coefficient measured? This is an easy question. The units of measurement have to be the same on both sides of the equation. That's clear. You can't say that 10 kilograms is more than 5 metres. The only comparisons can be between kilograms and kilograms, amperes and amperes, ergs and ergs. And so the units are the same on both sides of the equation.

Now, in the left-hand member of this equation we have a dimensionless number. By saying that the portion of absorbed radiation is equal to 1/30 or 0.08, we have said all there is to say. The units of measurement cancel out when intensity is divided by intensity. And that means that the quantity on the right must be nondimensional. Since thickness is measured in centimetres (or other units of length), the absorption coefficient is measured in inverse centimetres, that is, cm^{-1}.

Suppose a ray passes through a 10-cm-thick plate and loses 1% of its intensity. The left-hand side of the equation is equal to 1/100. This means the absorption coefficient here is equal to 0.001 cm^{-1}. Now if the radiation is soft and loses one percent of its energy when passing through a foil one micrometre (0.0001 cm) in thickness, then the absorption coefficient will be 100 cm^{-1}.

Physicists do not have a good theory for establishing a formula for the absorption coefficient. I can only say that the absorption coefficient is roughly proportional to the cube of the wavelength of X-radiation and to the cube of the atomic number of the substance through which the radiation is passing.

Since the wavelengths of X rays are extremely short, the frequencies of the electromagnetic waves are high. This means that a quantum of X rays, $h\nu$, carries a tremendous energy. This energy is only sufficient for chemical reactions leading to blackening of the emulsion of a photographic plate and the phosphorescence of screens (even light waves can do that), but there is even more than enough to break up molecules. In other words, X rays ionize air and other media through which they pass.

Now a few words about gamma rays. This term applies to the short-wave radiation that originates in radioactive decay. Gamma rays are emitted by naturally radioactive substances and are produced by artificial elements. A nuclear reactor generates gamma radiation of course. Very hard (high-energy) gamma rays are produced in the explosion of an atomic bomb.

Since gamma rays can have a very short wavelength, their absorption coefficient can be very small. For instance, the gamma rays that are emitted in the disintegration of radioactive cobalt are capable of passing through tens of centimetres of steel.

Substantial doses of short-wave electromagnetic radiation (such that is capable of disrupting molecules) are very dangerous to any living organism. Therefore protection is needed when dealing with X rays and gamma rays. Ordinarily it is in the form of lead. The walls of X-ray rooms are covered with a special coating containing barium salts.

Gamma rays, like X rays, can be used for radioscopy.

Usually use is made of the gamma rays of radioactive substances which are the "ashes" of nuclear fuel. Their advantage over X rays is the extreme penetrating power, but most important of all is the possibility of using, as the source of radiation, a small ampoule that can be fitted into places that are inaccessible to an X-ray tube.

X-ray Diffraction Analysis

In 1912 Roentgen was head of the chair of physics at the University of Munich. Problems dealing with the nature of X rays were constantly being discussed. I must say that Roentgen, who was an experimental physicist himself, had much respect for theory. There were a good many talented theoretical physicists at the chair of physics at the University of Munich that were cudgelling their brains over the nature of these new X rays.

And of course attempts were made to investigate the nature of the X rays by passing them through a diffraction grating. Using a diffraction grating, one can prove quite definitely the wave nature of light and also obtain exact determinations of the wavelength of one or another type of radiation.

One way of making such a grating is to rule fine scratches on a glass plate coated with aluminium; this is done with a cutting tool made of ivory. The lines thus cut must be strictly equally spaced. A good grating must have a small period and a large number of lines. Up to hundreds of thousands have been ruled and over a thousand per millimetre.

With the help of a lens, a strong point source of light can produce a parallel beam of light that falls on the grating at right angles. The rays emerge from each aperture and proceed in all directions (in other words, each aperture becomes the source of a spherical wave). But only in certain directions will the waves from all apertures be in-phase. For reinforcement, it is necessary that the path difference be equal to an integral number of wavelengths. Strong rays will go in directions at an angle α that obey the condition

$$a \sin \alpha = n\lambda$$

where n is an integer and a is the distance between slits. The reader can easily derive this formula without our help.

The whole number n is termed the *order of the spectrum*. If monochromatic light is incident on the grating, we obtain, in the focal plane of the eyepiece, several lines separated by dark intervals. If the light consists of waves of different wavelength, the grating will produce several spectra: of first, second, etc. orders. Each succeeding spectrum will be more extended than the preceding spectrum.

Since the wavelength of the light is of the same order as the distance between the slits, diffraction gratings disperse light (and not only visible light but also ultraviolet radiation and, best of all, infrared radiation) into spectra. These instruments make it possible to carry out a detailed spectral analysis of the radiation.

But in relation to X rays, diffraction gratings behaved like a system of open doors. X rays passed through them without deviating. It could be suspected that X rays are a flux of particles. But, on the other hand, there was every reason to think that this X-radiation is the same electromagnetic field as light, only with the wavelength λ much shorter. True enough, suppose the wavelength is very short. If that is so, then, according to the diffraction equation, all n rays proceeding at diffraction angles α from the line optical grating $a \sin \alpha = n\lambda$ will practically merge and no diffraction will be apparent. But to make a diffraction grating with slits separated by a distance a equal to millionths of a micrometre is out of the question. So what have we?

A young German physicist Max von Laue (1879-1960) was sure that X rays are a type of electromagnetic radiation. His acquaintance was a crystallographer who was convinced that a crystal is a three-dimensional lattice of atoms. In one of their numerous talks on scientific topics, Laue got the idea of correlating his view of the nature of X rays with the concept of a crystal as a lattice. Laue's idea was this: suppose the distances between the atoms of a crystal and the wavelength of X rays were quantities of the same order.

Can a three-dimensional lattice take the place of a line grating of slits? It was not obvious at all, but Laue decided to try it. The first experiment was simplicity itself. A beam of X rays was diaphragmed. A large crystal was placed in the path of the rays, and next to the crystal a photographic plate. True, it was not at all clear where to put the plate since the crystal was not a line lattice. The first attempts were thus failures. But soon, due to a mistake, they got the plate in the proper position.

This accidental mistake did not play any special role in the discovery because Laue was also working on the theory of the phenomenon besides trying to detect it. He was soon able to extend the theory of the line diffraction grating to the three-dimensional case. From the theory it followed that diffracted rays would appear only in the case of certain orientations of the crystal with respect to the incident rays. From the theory it also followed that the most intensive rays were those deflected at small angles. And from this it was evident that the photographic plate had to be positioned behind the crystal perpendicular to the incident rays.

Among the first researchers to notice this discovery were the Braggs, the father Sir William Henry (1862-1942) and his son Sir William Lawrence (1890-), English physicists. They repeated Laue's experiment, provided a very simple and pictorial interpretation and also demonstrated many simple instances in which Laue's discovery could be used as a method for studying the atomic structure of substances.

Let us take a brief look at *X-ray diffraction analysis* and also examine the path by which we can determine the structure of a crystal, that is, measure the distances between atoms with an accuracy of up to one hundredth of an angstrom and give a picture of the spatial arrangement of atoms in a molecule, and also determine the nature of the packing of molecules in a crystal.

Suppose a crystal is set in a special holder and rotates on an axis. The X ray falls at right angles to the axis of rotation. What happens? Let us examine the diffraction phenomena that take place when an X ray is incident on the crystal so that a lattice point is regarded as the scattering centre.

The Braggs (father and son) demonstrated that the scattering of X rays by lattice points is equivalent to a curious kind of selective (that is, taking place only for certain discrete values of the angle) reflection of rays from a system of nodal planes, into which the lattice may be partitioned.

Suppose a ray, which is an electromagnetic wave of a definite length, falls on a crystal at some angle. This angle will differ for different systems of planes. We have every right to assume that any atomic plane will reflect the X ray in accord with the law: the angle of incidence equals the angle of reflection. But there is essential difference from optical rays. Unlike light, X-radiation penetrates deep into a crystal. This means that the reflection of the ray will occur not from the external surface but from all atomic planes.

Let us consider one such system characterized by the interplanar spacing d. Each of them will "reflect" an incident ray at one and the same angle θ. These reflected rays must interfere, and a strong secondary ray will arise

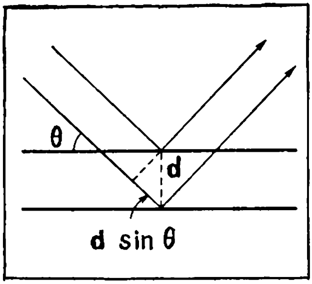


Figure 3.2

only if the rays reflected from all planes of the family are in the same phase. In other words, the path difference between the rays must be equal to a whole number of wavelengths.

Figure 3.2 depicts a geometric construction from which it follows that the path difference between adjacent reflected rays is equal to $2d \sin \theta$. Hence the condition of diffraction will be of the form

$$2d \sin \theta = n\lambda$$

The Russian crystallographer G. V Vulf arrived at this formula at the same time as the Braggs and so it is called the *Bragg-Vulf equation* (also known as *Bragg's law* or *Bragg's equation*).

A crystal can be partitioned into systems of planes in numberless ways. But only a certain system becomes effective for reflection: that in which the interplanar spacing and orientation with respect to the incident ray are such that Bragg's equation is obeyed.

Obviously, if the ray is monochromatic (that is, the electromagnetic wave has a definite length), then reflection may not occur in the case of an arbitrary position of the crystal with respect to the ray. However, by rotating the crystal we can bring different systems of planes into the reflecting position in turn. This appeared to be the most suitable way for practical work.

As for Laue's experiment, its success stemmed from the fact that the radiation incident on the crystal was the "white spectrum" of X rays, that is, a flux of waves whose lengths are continuously distributed over a certain interval (see below). Therefore, although the crystal was fixed in Laue's experiment, different systems of planes appeared in the "reflecting" position for waves of different length.

At the present time, X-ray diffraction analysis is completely automatic. A small crystal (0.1 to 1 mm) is fixed in a special holder, which can turn the crystal according to a specified programme, bringing into the reflecting position all its system of planes one after the other. Every reflecting plane (we'll use that word so as not to repeat "system" all the time) is characterized, first, by its interplanar spacing, second, by the angles which it forms with the axes of an elementary cell of the crystal (lengths of the edges and the angles between the edges of the cell are measured first of all and automatically) and, third, by the intensity of the reflected ray.

The more atoms a molecule contains the greater (naturally) the dimensions of the elementary cell. This increased complexity brings with it a greater volume of information. The point is that the larger the cell the greater the number of reflecting planes. The number of measured reflections can range from several tens to several thousands.

We promised a brief survey of the idea underlying X-ray diffraction analysis. Let us first turn the problem around. Suppose we know everything about the structure of a crystal. This means we know the pattern of atoms, that is, we know the coordinates of all atoms that make up an elementary cell (I suggest the reader go back to the second book of this series and refresh his memory concerning crystal structure). Let us consider some kind of system of reflecting planes. Obviously the following is sufficient. If most of the atoms of the crystal lie in planes passing through the lattice points, then all atoms will scatter the X rays in a single phase. What we get is a strong reflected ray. Now take another case. Half of the atoms lie in the nodal planes, and the other half lie in between the reflecting planes. Then half of the atoms scatter incident light in one phase while the other half scatter it in the opposite phase. There will be no reflection!

These are two extreme cases. In all other cases we obtain rays of different intensity. A measuring instrument (called an *automatic diffractometer*) is capable of measuring the intensities of reflections that differ by a factor of ten thousand.

The intensity of a ray is uniquely connected with the arrangement of the atoms between the nodal planes. The formula stating this relationship is too unwieldy to give and, what is more, we don't even need it. What we have already described about the two extreme cases should be enough for the reader to be convinced there is a formula in which intensity is represented as a function of the coordinates of all atoms. The type of atoms is also taken into account in this formula because the more electrons there are in an atom the more intensely it scatters X rays.

Also included in the formula relating structure and intensity of reflected rays is of course information about the orientation of the reflecting plane and also about the dimensions of the elementary cell. We can write down as many such equations as there are measured reflections.

If the structure is known, then the intensities of all rays can be computed and correlated with experiment. But that isn't the problem that confronts us! What we have before us is the converse problem: using information concerning the intensities of several tens or hundreds or thousands of reflections, find the coordinates of all atoms in a cell. At first glance it might seem that with the tremendous calculating power of modern computers the converse problem can be solved with ease. A lot of equations? But computers can surely handle them.

However, the matter is not at all so simple. Experimental findings have to do with the intensities of the rays. The intensity is proportional to the square of the amplitude. The formula discussed above is actually a formula of interference. The waves scattered by all atoms of a crystal interfere with one another. What we have is a composition of amplitudes scattered by all atoms. The overall amplitude is computed and the intensity is found by squaring the amplitude. To solve that problem is easy. But how do we handle the converse problem? Take the square root of the intensity in order to obtain the amplitude? Correct. But a square root has two signs.

I just hope the complexity of the problem is now clear to the reader. We have equations enough (more than enough) to find the coordinates of the atoms. But in the right-hand member of the equation we have numbers that are known only up to sign.

We are thus at an impasse. And, at first, researchers did not attempt to solve the converse problem. They employed the trial and error method. They assumed, on the basis of information concerning allied structures, that the unknown structure was such and such. Then they calculated the intensities of a dozen rays and compared them with experimental findings. And failed. And then they took up another model of structure.

In simple cases this approach produced correct results, though there were difficulties. But when the "structure men" (which is what these researchers were being called) had studied nearly all the simple structures, they came to a halt when confronted with the converse problem.

In the middle of the 1930s it was conjectured that even complex structures might be "solved" (that was the lingo then) if one confined himself to a study of molecules that contain many light atoms and one heavy atom. A heavy atom contains many electrons and scatters X rays much more strongly than light atoms. Therefore, to a first approximation (very rough) it may be assumed that the crystal consists solely of heavy atoms. If there is one atom in a cell, its coordinates can be found with relative ease by means of trial and error. With these coordinates, it was then assumed that only that atom functioned in the crystal, and it was further assumed that the signs of the amplitudes determined for the fictitious structure consisting of only heavy atoms were the same as for the actual structure.

A highly important discovery was made some twenty years ago: namely, proof was given of a theorem concerning the relationship between the amplitudes of reflections of different families of planes. For example, a relationship was found between the signs of the amplitudes of three reflections that are shifted in phase with respect to the point of the cell by α, β, and $\alpha+\beta$. It turns out that if the product $\cos\alpha \cos\beta \cos(\alpha+\beta)$ exceeds 1/8 in absolute value, then it must have a positive sign. You can check for yourself if you don't believe me.

The development of that idea led to what are known as direct methods of structure analysis. Even in rather complicated cases the experimental device can be hooked up to a computer, which will then "spew out" the structure of the crystal.

When the signs of the amplitudes of reflection have been established, then determining the coordinates of the atoms becomes, as we now know, a problem involving the solution of a very large number of equations with many unknowns. It is important here that the number of equations be at least ten (better, a hundred) times the number of desired coordinates of the atoms.

I will not go into the techniques used to solve such systems of equations. Actually, the problem is circum-

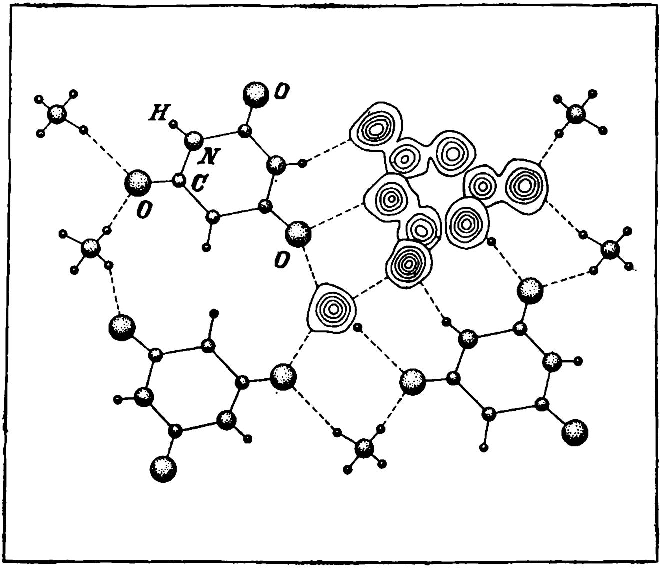


Figure 3.3

vented and is reduced to the construction of what are known as Fourier series of electron density. Only a reader with a good mathematical background could stomach any discussion of the theory of Fourier series or, still worse, its application to the problem of determining structure. But we won't need such knowledge here. The main thing is that the idea as such has been discussed. I think that is sufficient for our purpose.

In what form does the physicist (the specialist in X-ray diffraction analysis) provide the chemist with information concerning structure? Some idea of this can be gained by a glance at Figure 3.3, which shows a very simple structure called ammonium barbiturate. Today, determining the structure of such complexity is child's play. An automatic device gives the structure without any effort on the part of the researcher. The electronic computer can print out the numbers (the values of the coordinates of the atoms) or can produce displays like that of Figure 3.3. Atoms of different kinds are symbolized by circles of different size. And if the researcher wants the computer to give a picture of electron density, it can do so. Each atom is displayed in the way that geographers portray lines of equal altitude in mountainous regions. But in our case the closed circuits are not lines of equal altitude but curves indicating the electron density at every point. The tip of the "mountain peak" is the centre of the atom.

The accompanying diagram is a very minute portion of the contribution that this method has made to science. The success of the method is colossal. To date, the structures of over 15 thousand crystals have been deciphered, and these include several tens of structures of proteins whose molecules consist of many thousands of atoms.

Determining the structures of complex molecules lays the foundation of biological chemistry and biological physics, two new sciences that have burst into bloom and are expected to give us discoveries of the secrets of life, illness, and death. Despite its sixty-some years of existence, X-ray diffraction analysis remains at the forefront of science.

The X-ray Spectrum

In the preceding section we mentioned the existence of a "white" spectrum and monochromatic rays. How can we decipher the nature of the spectrum of hard electromagnetic radiation? And when is it "white" and in what cases is it monochromatic?

If we diaphragm X rays or gamma rays coming from some source (that is, if we place in their path two screens

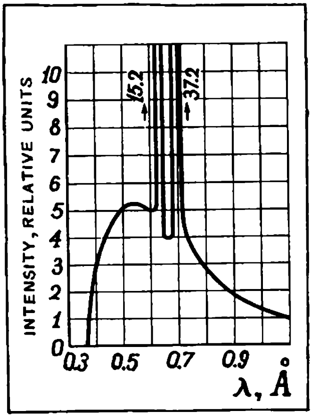


Figure 3.4

with small openings) and make the beam impinge on a crystal, then in the most general case we obtain several rays reflected from planes that are in positions that satisfy the Bragg-Vulf equation. If we position the crystal so that some plane (yielding a strong reflection) coincides with the axis of rotation of a special instrument (called an *X-ray spectrograph*) and then turn the crystal so that the plane comes under the incident ray at a succession of all angles θ, then a component of the spectrum with a definite wavelength will be reflected for each position of the crystal. This reflected wave can be either picked up by an ionization counter or recorded on film. In this way we can create a monochromatic ray of any wavelength of the radiation spectrum and, also, develop a method for exploring the spectrum of any type of radiation.

A typical spectrum of an X-ray tube with molybdenum anode is shown in Figure 3.4 (at 35 kilovolts). The conclusion is almost immediate that there are two reasons for the generation of an X-ray spectrum. We see that the observed spectrum is a superposition of sharp peaks on a continuous curve. Of course, the origin of these peaks differs from the origin of the solid curve.

Immediately after the discovery of the diffraction of X rays, a study began of X-ray spectra and the following was established. A continuous spectrum is not typical of the material of the anode and depends on the voltage. Its peculiarity is that it breaks off sharply at a certain minimum wavelength. When moving in the direction of long wavelengths, the curve passes its maximum and smoothly falls off without having any apparent endpoint.

By increasing the voltage in the X-ray tube, researchers demonstrated that the intensity of the continuous spectrum grows and the boundary is shifted towards the short waves. And the following very simple equation was established for the boundary wavelength:

$$\lambda_{\text{min}} = \frac{12.34}{U}$$

In quantum language, the statement of the rule is simple. The quantity eU is the energy gained by an electron in covering the distance from the cathode to the anode. Naturally, the electron cannot release more energy than this quantity indicates. If it transfers all the energy to generating an X-ray quantum ($eU = h\nu$), then after substituting the constants we obtain the above equation (λ is in angstroms and U in kilovolts).

Since we get a continuous spectrum, it follows that the electrons do not necessarily give up all their energy to the generation of X rays. Experiment shows that a large part of the energy of an electron beam is converted into heat. The efficiency of an X-ray tube is very low. The anode heats up and it has to be cooled by means of water flowing inside the anode.

Is there a theory that can explain the reason for a continuous spectrum of X rays? Yes, there is. Calculations (which unfortunately we cannot give here) show that from the general laws of the electromagnetic field (from Maxwell's equations) that were discussed in book three of this series, it follows very rigorously that if electrons are decelerated, the result is a continuous spectrum of X rays. Collisions with a solid are not essential at all. If the electrons are decelerated by means of a counter-field, the result is a continuous emission of X rays without any participation of the anode in the game.

There is yet another way of encountering a continuous X-ray spectrum. We recall that a continuous electromagnetic spectrum is emitted by incandescent bodies. In terrestrial conditions we do not encounter an X-ray spectrum of such origin because (compare the formula given on page 15) at the highest possible temperature of an incandescent body (6000 degrees—since no solid body can stand up to a higher temperature) the wavelength of thermal radiation will be close to half a micrometre. But we must not forget about the existence of plasma. In stellar interiors and in artificial plasma produced here on earth, temperatures in the millions of degrees can be obtained. Then the thermal spectrum of electromagnetic radiation will also include X rays. X rays coming from outer space help resolve intriguing problems of astrophysics.

Now let us touch on the sharp peaks that are superimposed on the curve of the continuous spectrum.

Just the reverse was proved with respect to these rays, the reverse with respect to the law of the continuous spectrum. The sites of the peaks, that is, their wavelengths, are unambiguously determined by the material of the anode. Such emission therefore goes by the name *characteristic*.

Its origin is unbiasedly accounted for by the quantum model of the atom. The electron rays of an X-ray tube are capable of penetrating an atom of the anode and knocking out electrons from the very lowest energy levels. As soon as a low level is vacated, that site is filled by an electron more distant from the centre of the atom. Energy is emitted in accordance with the basic quantum law: $E_m - E_n = h\nu$. Energy levels are arranged in different ways in different atoms. It is therefore natural that the new spectra will be characteristic spectra.

Since the lines of a characteristic spectrum are the strongest ones, they are used for X-ray diffraction analysis. It is best to eliminate the continuous spectrum; that is, before making the ray fall on the crystal under investigation, it has to be reflected from the crystal monochromator.

Since the spectra of different elements are characteristic, a dispersion of the ray into a spectrum may be used for purposes of chemical analysis. It is called *X-ray spectral analysis*. There is a whole range of fields (for example the study of rare-earth elements) where X-ray spectral analysis is absolutely indispensable. The intensities of spectral X-ray characteristic lines make it possible to determine to a high degree of accuracy the percentage content of any element in a mixture.

Now a word or two concerning the spectra of gamma rays. In terrestrial conditions, we have to do with gamma rays that originate in radioactive decay (this will be discussed later on). Radioactive disintegration may or may not be accompanied by the emission of gamma radiation. But no matter what the type of radioactive decay, the spectrum of gamma radiation will be characteristic.

If characteristic X rays appear when an atom falls from a high energy level to a lower level, then gamma rays are produced as a result of a similar transition of the atomic nucleus.

The gamma-ray spectra of radioactive transformations have been thoroughly studied. There are tables in which one can find exact data on the wavelengths of gamma rays generated in alpha- and beta-transformations of one or another radioactive isotope.

Radiography of Materials

I have time and again had to call attention to the fact that terminology is a weak spot in science. Science develops at such a rate that the meaning of a term changes in the course of a single generation. And at the same time, changes in terminology are associated with alterations of what is taken to be accepted. What is more, old books cannot be taken out of circulation, and so the only thing left is to include strict definitions of every term that is employed and to redefine others.

At the present time, when one speaks of X-ray diffraction analysis, he has in view the investigation of the atomic structure of crystals. The object of study is a single crystal of the substance.

However, studying structure by means of X rays does not in any way exhaust the problem. Characteristic and information-rich pictures are obtained if radiographs are taken of any materials and not only single crystals. In such cases we ordinarily use the term *radiography*.

If a piece of metal foil is placed in the path of a monochromatic X-ray beam, a system of concentric circles appears on a flat photographic plate. A radiograph of this type is termed a *Debye crystallogram*. What is its origin?

Most solids consist of tiny crystals oriented at random with respect to one another. When a melt of some substance begins to cool, crystallization sets in simultaneously at a large number of sites. Each minute crystal grows every which way and the growth continues until the crystals meet.

Each crystal has the same system of atomic planes; this is because we are dealing with structurally identical crystals. Let us examine one of these systems of planes with interplanar spacing *d*. There are a large number of minute crystals (ordinarily, the linear dimensions of a solid-body crystal is, as to order of magnitude, equal

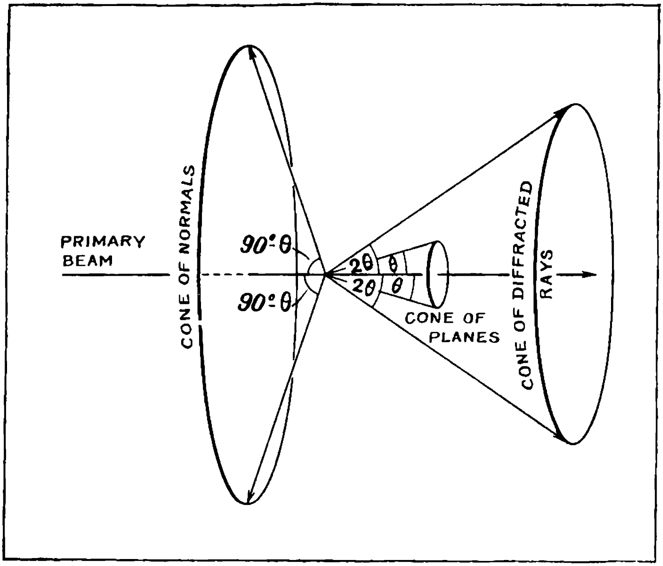


Figure 3.5

to one ten-thousandth of a centimetre), and among them are those whose planes lie at an angle θ to the incident ray, which angle satisfies the Bragg-Vulf equation. Each such crystal will leave a spot on a photographic plate. Reflections will be produced by all such minute crystals whose normals to the planes form a cone (Figure 3.5). If that is so, then the reflected rays will lie in the cone. The intersection of that cone with the photographic plate yields a circle. By measuring the radii of the circles and knowing the distance from the object to the photographic plate, we can find the Bragg angle θ at once and will then be able to calculate all the interplanar spacings of the substance on the basis of such a radiograph.

Using such a diffraction pattern, we can at once distinguish any amorphous material from crystalline material. Amorphous bodies do not have reflecting planes. Therefore the radiograph will not reveal any system of distinct diffraction rings. There is always some sort of order in the arrangement of molecules of a substance for the simple reason that atoms cannot "overlap" in any way. And as calculations show, this results in the radiograph of an amorphous body having one or (rarely) two smeared rings.

However, the rings observed on radiographs yield a good deal of valuable information concerning the structure of materials such as metals, polymers, and natural compounds. If the substance consists of large crystallites, then the diffraction ring will not be continuous and will consist of separate small spots. If the crystallites are not arranged haphazardly but are oriented along an axis or plane, as occurs in wire or sheets of metal, in polymer strands and in vegetable fibres, then the diffraction rings will tell us so immediately. It is easy to see that if we have preferential orientations of crystallites, the reflections from atomic planes will not fill the cone of rays in a continuous fashion. In place of rings, the radiograph reveals arcs. And if the orientation is highly refined, these arcs may degenerate into tiny spots.

Of course, to give a detailed description of the type of structure from the form of radiograph is no easy problem. In this case, too, the method of trial and error plays an essential role. The investigator thinks up models of the structure of materials, calculates patterns of reflections of X rays that should yield the models he has concocted and, by comparing calculations and experiment, chooses the proper pattern for the structure of the substance.

In the radiography of materials we distinguish, somewhat artificially, between scattering through large and small angles. From the Bragg-Vulf equation given above, it is clear that scattering through large angles occurs if a structural periodicity is observed at small distances (say, 3 to 10 Å). If the reflected (we could say, scattered) X rays yield a diffraction pattern that collects near the primary beam, then this means the structure possesses periodicity at large distances.

In the science of metals we deal mainly with diffraction rings located at large angles, since they consist of crystallites whose atoms form regular lattices with cells having dimensions of the order of units of angstroms.

When the objects of investigation are substances made up of macromolecules, we encounter a very interesting situation. (By the way, macromolecules are found in many natural substances such as cellulose or DNA and also synthetic polymer materials with such familiar names as polyethylene, nylon, capron and so on.) In some cases we obtain radiographs exhibiting rings of large diameter only. In other words, we have to do with the same kind of scattering through large angles as is the case in metals. And at other times, on occasion, we do not find large-diameter rings but rather diffraction rays that only slightly deviate from the primary direction. Finally, there are cases where the substance reveals X-ray scattering both through large and small angles.

Small-angle scattering (of course the division into types of small-angle and large-angle scattering is somewhat artificial) is in the range from several minutes to 3-4°. Naturally, the smaller the diffraction angle the greater the repeating period of structural elements that produced the diffraction.

Large-angle scattering is due to the order in the location of atoms inside the crystallites. Small-angle scattering, on the other hand, is connected with the ordered arrangement of rather large formations that are called permolecular. It can even turn out that there is no order at all inside these formations consisting of hundreds and even thousands of atoms. But if such large systems form one-dimensional, two-dimensional or three-dimensional lattices, then small-angle X-ray scattering will tell the story. To get a picture of what I mean, imagine a carefully arranged structure made up of bags of potatoes. It is extremely interesting, and probably there is profound meaning in the fact that we encounter just such a "baggy" order in a very large number of biological systems. For example, the long molecules that make up muscle tissue are arranged very accurately, like a package of round pencils. As small-angle X-ray scattering shows, we find just such extraordinarily high order in cell membranes, in such protein systems as viruses. and so on.

There is an interesting theorem in the theory of diffraction. I will not try to prove it but I'm sure it will appear natural to the reader. It can be rigorously demonstrated that the type of diffraction pattern remains the same if in the object that produces the diffraction we interchange the apertures and opaque places. This theorem sometimes makes the research worker miserable. It is when X-ray scattering can be caused either by pores within the substance or by extraneous inclusions. The study of pores—their dimensions, shape, and quantity per unit volume—is of great practical interest. The colouring of synthetic fibres depends to a very large extent on these peculiarities in their structure. It is easy to see that an uneven distribution of pores will result in an uneven colouring and a final unpleasant fabric.

From what has been said it is now sufficiently obvious that radiography of materials is not only a method of studying substances but also a method of technical control in a great variety of industries.

4. Generalizations of Mechanics

Relativistic Mechanics

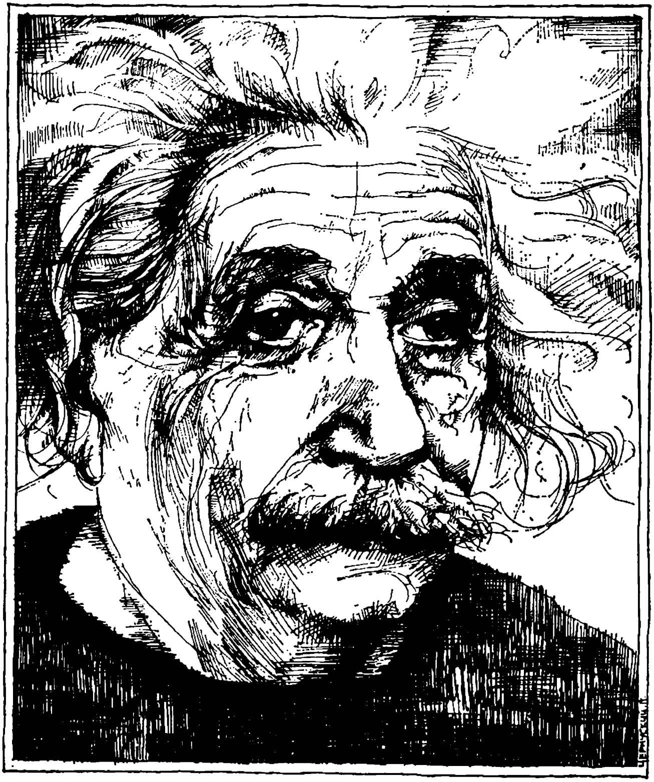
Newtonian mechanics that we discussed in the first book of this series is one of the greatest achievements of the human genius. It made it possible to compute the paths of planets, the trajectories of rockets and spacecraft, the behaviour of mechanisms. The development of physics in the twentieth century demonstrated that the laws of Newtonian mechanics have two limitations: they become unsatisfactory when dealing with the motion of particles of small mass and they fail us when we have to do with the motion of bodies with velocities close to the velocity of light. For small particles, Newtonian mechanics is replaced by the so-called *wave mechanics*; for fast bodies, it gives way to *relativistic mechanics*.

Classical mechanics is also made more complicated when we encounter very great forces of gravitation. The unimaginably strong fields of gravitation that govern the behaviour of superdense stars force us beyond the limitations of such simple formulas of mechanics as were explained in the first book. We will not dwell on such changes here but will take up the two highly important generalizations just mentioned when dealing with the motions of microparticles and when studying velocities comparable to the velocity of light.

Let us begin with relativistic mechanics. The path leading to this important division of physics least of all resembles a straight road. Not only is it tortuous but it even passes through quite different countries. The starting point here is the ether. Generally speaking, physicists were having it easy at the end of the 19th century. Max Planck's teacher did not advise him to take up physics because that science, he said, is practically completed. There were only two slight blemishes on the otherwise perfect edifice of physics: drawbacks in accounting for the radiation of a black body (when that was cleared up, physicists discovered the quantum) and the distressing experiment of Michelson. This experiment, which proved that the velocity of light does not combine with the velocity of the earth in its orbital motion and is the same in all directions, caused some anxiety about the properties of the ether.

Hardly anyone doubted the existence of some kind of delicate matter whose oscillations constituted the electromagnetic waves. After over a hundred years have passed it seems even surprising that despite the great number of inconsistencies that the "ether" hypothesis led to, the vast majority of scientists—some extremely talented—went to great lengths to introduce no end of supplementary assumptions with the sole purpose of salvaging the concept of light as the motion of an invisible ultimate material substance.

Some pictured the ether as a placid sea through which the planets moved; others thought that the ether could be entrained like air by moving bodies. As strange as it may now seem, nobody suggested the obvious idea that the oscillations of electric and magnetic vectors occur *at a point*, and for that reason cannot be accounted for in terms of mechanical displacements. Every effort was made to hold onto the old concepts, theories were set forth in which formally proper mathematical expressions were derived (the infamous square root $\sqrt{1-(v/c)^2}$ included as well; here v was the velocity of a body and c the velocity of light), but the equations were wrongly inter-



Albert Einstein (1879-1955)—the genius who created the theory of relativity and revolutionized all physical thinking. In 1905, Einstein published a treatise devoted to the special theory of relativity. In 1907, he obtained a formula relating energy and the mass of a body. In 1915, Einstein published his general theory of relativity. From this theory there followed new laws of gravitation and conclusions concerning the curvature of space.

preted. Particularly upsetting was the Michelson experiment, first carried out in 1881. Using an interferometer (described in Chapter 2), Michelson demonstrated that the velocities of light along and across the orbital motion of the earth are practically the same.

Even this fact (completely upsetting the theory concerning the existence of an ether) did not compel leading physicists to give up faith in a tenuous matter permeating all bodies. Michelson's experiment forces us to give up any idea of an ethereal wind? Good and well, we will give up the wind and the universe will be still more beautiful with a motionless ether, and we will accept the Newtonian absolute space with respect to which all celestial bodies move.

To account for the Michelson experiment, such outstanding physicists as Sir Joseph Larmor (1857-1942) and Hendrik Antoon Lorentz (1853-1928) applied the hypothesis of the contraction of bodies in the direction of their motion. However, logical inconsistencies and the artificial nature of the explanations of many phenomena concerning electrodynamics continued to leave a sense of inadequacy.

The Gordian knot of contradictions was finally cut by the greatest physicist of our century, Albert Einstein (1879-1955).

The starting point in Einstein's reasoning was the principle of relativity. There was hardly any doubt in anyone's mind, after Galileo, that in regard to mechanical motions, all inertial systems are equivalent (please

---

The theory of relativity does not exhaust Einstein's contribution to physics. From the research of Max Planck he drew the conclusion that there exists a light particle (called a photon) and demonstrated that from this standpoint it is possible to account for a number of fundamental phenomena, including the photoeffect.

turn back to book one in this series and review the facts about inertial systems). What appeared to be rather strange and esthetically unpleasant was that we have equivalence for mechanical motion but not for electromagnetic phenomena. Now suppose we assume the principle of relativity to hold true for all phenomena. But doesn't that lead to a contradiction? Imagine a source emitting a light ray. The light wave sets out into space and, observing it, I permit myself to disregard the fact that the source may be in motion or at rest. But if that is so, then the velocity of electromagnetic waves (299 792 km/s) must be the same from the viewpoint of observers living in systems moving with respect to one another with an arbitrarily high velocity v.

Suppose we have a train on a straight section moving at a constant speed of v. Parallel to the track is a highway and on it and moving in the same direction is a speeding motorcyclist. A traffic officer sees the culprit and with his miniature radar quickly finds that the motorcycle is doing 85 km/hr. The locomotive engineer glances at the motorcyclist and sees him catching up with the train and then overtaking it. This observer has no difficulty in measuring the speed of the motorcycle. It is $u'=35$ km/hr. It is obvious that the speed of the train is 50 km/hr. The law of composition of velocities holds true:

$$u = v + u'$$

This most obvious of all rules does not fit the case of light rays. Photons move with the same velocity with respect to two observers located in different inertial systems.

The genius of Einstein lay in the fact that he rejected this most obvious conclusion not only for light but, aiming to retain a unified approach to all physical phenomena (both electromagnetic and mechanical), was so bold as to give up the law of composition of velocities for all bodies.

From this standpoint, there is no need to explain anything in Michelson's experiment. Since the velocity of light does not depend on the velocity of motion of the source, it (the velocity) will be the same in all directions: both along the earth's orbit and across the orbital path of the earth around the sun.

It follows immediately from the principles just formulated that the velocity of light is the maximum velocity.* Indeed, if the velocity of light is never combined with the velocity of any source, that means it is impossible to overtake light. In his reminiscences, Einstein recalls that as early as 1896 he was considering the following question: "If it were possible to set out after a light wave with the velocity of light, then would we be dealing with a time-independent wave field? That does indeed seem to be impossible."

To summarize, then: there is no particle that can move with a velocity greater than the velocity of light. Let us give some thought to this statement. It is so paradoxical that it needs repeating. If an electromagnetic wave takes off from the earth or another planet in some direction, then the rate of propagation of that wave, when measured by a terrestrial observer and by an observer flying over the earth in a spacecraft moving with a fantastic

*Generally speaking, however, relativistic mechanics is not contradicted by the existence of particles moving with velocities arbitrarily greater than the velocity of light. Theoreticians have even given them a name: they are called *tachyons*. However, if such particles existed, the velocity of light would still be a limiting velocity for them—it would be the minimum velocity not the maximum velocity. I personally think that the theory of tachyons is nothing more than an elegant mathematical toy. If the world of tachyons did indeed exist, then events taking place in the universe would, in principle, never be influenced by that world.

velocity, is the same. This assertion holds true also for any particle moving with a velocity equal to the velocity of electromagnetic waves. Light is no exception in the Einstein theory.

Now what happens when the velocity of a moving body is less than the velocity of light? Obviously, in this case too the simple principle of the composition of velocities that is so usual does not hold true. However, the deviation from the ordinary rule for composition of velocities begins to be felt only when the velocity of the body is very great.

Relativistic mechanics—that is the name given to the mechanics of fast-moving bodies—leads to the following rule for the composition of velocities:

$$V = \frac{v + v'}{1 + \frac{vv'}{c^2}}$$

Make a rough estimate of what the values of v and v' must be for corrections to the simple rule for composition of velocities to be needed.

How about spaceflight? Does the ordinary rule for the composition of velocities function when we deal with speeds of tens of kilometres per second?

We know that a good way to gain speed is to launch a second rocket system from a spaceship in flight. This could be one way of launching spacecraft to the outer reaches of the solar system. Denote by v the velocity of the spaceship relative to the earth, and by v' the velocity of the craft launched from the spaceship relative to the spaceship. Suppose both velocities, v and v', are equal to 10 km/s. Now compute the exact velocity of the spacecraft relative to the earth, using the exact formula of the composition of velocities. To the 1 in the denominator we have to add the fraction $10^2/(9\times10^{10})=10^{-9}$. The correction is practically negligible, which means the classical rule for the composition of velocities holds true.

The question now is: When does relativistic mechanics come into its own? We will have the answer soon, but meanwhile let us continue the consequences that stem from the hypotheses just formulated. Since we are forced to give up the principle of composition of velocities, we must be ready to introduce essential corrections in the other formulas of mechanics as well.

As we have already mentioned, it was Michelson's experiment that played a decisive role in the development of the theory of relativity. Michelson demonstrated conclusively that the velocity of light along the earth's orbit and across it is the same.

We will not consider the path of rays in the Michelson interferometer and will confine ourselves to simpler events. Somewhere on the earth, a very simple laser experiment is set up on a tower at an altitude l above the earth's surface. The ultrafine laser beam goes in the direction of the earth's radius, is reflected from a mirror placed on the earth's surface, is returned to where it started from, and is received by a photocell that engineers place in such a position that we regard the light source and the receiver of light to be in a single point. In Figure 4.1 it is labelled S. Using an ultrarefined stopwatch, we can fix two instants: the first when the light set out and the second when it reached the photocell.

Two observers follow this event. The first one stands next to the experimental setup, the other one is perched on a distant star. Both observers measure the time interval τ between the two events: the emission of light and its return to point S. The first draws the path of the ray and it turns out to be simplicity itself. He believes that the paths of the ray there and back coincide absolutely. His reasoning is corroborated by the equation $\tau = 2l/c$.

The stellar observer follows the flash as the light starts out on its journey and records its arrival at the photocell.

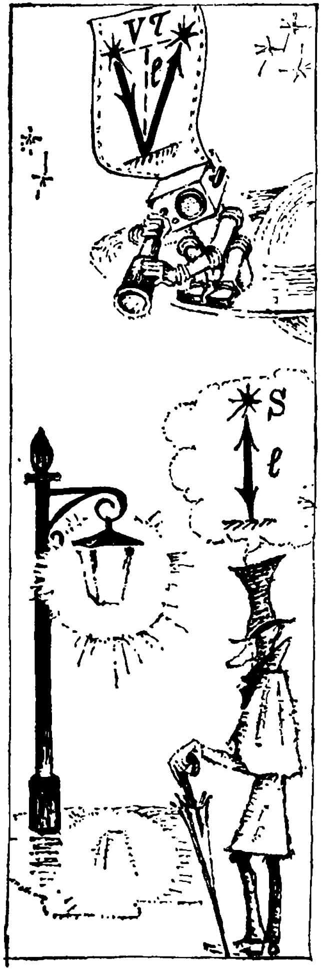


Figure 4.1

The time interval that he measures is equal to τ. To be sure everything is properly done, he too draws the path of the ray. For him, however, the positions of point S at the time the stopwatch was switched on and the instant he recorded the reaction of the photocell do not coincide. The path of the light ray in his case is different. The stellar observer knows the velocity of the earth relative to himself and so what he gets is an equilateral triangle, the base of which is equal to $v\tau$ and the altitude to l. Using the Pythagorean theorem, the stellar observer finds that the path traversed by the light ray is equal to $2\sqrt{l^2+(v\tau/2)^2}$. This path is equal to $c\tau$, because the velocity of light is the same for all observers. If that is so, the time interval between the two instants is equal to

$$\tau = \frac{2l}{c\sqrt{1-\frac{v^2}{c^2}}}$$

That is a surprising result! From the standpoint of the terrestrial observer, the same time interval between those very same events is equal to $2l/c$.

Appealing to logic, we can draw an unavoidable conclusion: the time reckoned by an observer at rest differs from the time reckoned by an observer in motion.

The time of the stationary observer is called the *proper time* and is denoted by τ_0. We find that the time of an observer moving with velocity v is connected with the proper time by the expression

$$\tau = \frac{\tau_0}{\sqrt{1-\beta^2}}, \text{ where } \beta = \frac{v}{c}$$

What this means is that a moving clock goes more slowly than a stationary clock. If we accept the basic postulates of the theory, we cannot escape such a conclusion. And this leads to the strange, at first glance, consequence that the idea of simultaneity has to be given up.

But won't it turn out, then, that from the viewpoint of one observer, Jim fired a gun and John was killed by the bullet while, from another point of view, John was killed first and Jim fired later? The reader can rest assured that relativistic mechanics does not allow for any such nonsense. The principle of causality will never be upset. And the proof can even be given popularly, but, unfortunately, that is beyond the scope of this book.

However, a few words are in order about the so-called twin paradox which even now is brought forth as proof of the inconsistency of the theory. John and Pete are twins. Pete takes leave of John and sets out on a space voyage with a velocity close to the velocity of light, and after a certain time returns to earth. Pete's clock has been going slower and so he returns to earth still fresh and young, whereas his brother John has grown old and feeble.

However—fortunately or unfortunately, depending on who likes what—no such meeting can be organized, that is, if we keep to the conditions under which the formulas under discussion hold true. The point is that Pete would have to reverse his velocity, and so the conclusions referring to inertial systems have nothing to do with this case.

The relativity of time is not the only consequence of the new theory. Just as the proper clock of the observer goes faster than any other clock, so also the length of a rod, l_0, held in one's hand is the maximum length. From the point of view of any observer moving with velocity v along the rod, that same length is equal to

$$l_0\sqrt{1-\beta^2}.$$

The same root also appears in the expression for the mass. The mass m_0 of the body that the observer holds in his hands is termed the *rest mass*. It is minimum. For a moving observer,

$$m = \frac{m_0}{\sqrt{1-\beta^2}}$$

It is quite natural that the mass increases with an increase in velocity. Indeed, if velocity has a limit, then as a particle approaches that limit, it becomes progressively more difficult to accelerate the moving particle. That is precisely what is meant when we say that the mass of the particle increases.

For a long time, no one encountered such high velocities that would force one to take into account the difference between the square root and unity in the formulas for distance and time. It was only just recently that the truth of the formula for time was demonstrated.

Now about the dependence of the mass on the velocity, it was revealed in the case of a stream of electrons even before Einstein's article appeared. The formula for the mass is an engineering formula in the full sense of the word. As we shall see in the next section, without it one simply could not even design a modern particle accelerator. In these very expensive machines, nuclear particles are accelerated to such an extent that the square root gets much closer to zero than to unity.

The formula illustrating the dependence of mass on velocity was first proposed before Einstein. But before the advent of relativistic mechanics its interpretation was not correct.

However, the famous expression $E=mc^2$, which connects mass and energy, was derived by Einstein. That formula and also the dependence of l, τ, and m on the velocity follow rigorously from the postulates of the theory.

Multiply the mass by the square of the velocity of light. For a moving body it is mc^2, for the same body at rest it is m_0c^2. Let us form the difference of these two expressions:

$$mc^2 - m_0c^2 = m_0c^2 \left(\frac{1}{\sqrt{1-\beta^2}} - 1\right)$$

Taking advantage of the approximate equation, the truth of which you can verify with ease, we have

$$\frac{1}{\sqrt{1-\beta^2}} = 1 + \frac{1}{2}\beta^2$$

The difference that we are calculating is of the form

$$mc^2 - m_0c^2 = \frac{m_0v^2}{2}$$

As you can see, it is equal to the kinetic energy of the body.

Thinking about this equation, Einstein came to the following fundamental conclusion. The energy of a moving body may be expressed as

$$E = mc^2$$

This energy is made up of the rest energy m_0c^2 and the energy of motion. Without having any information about the structure of a body and not knowing the nature of the interactions of its component particles, we can state that its internal energy is equal to

$$U = m_0c^2$$

The internal energy of a body of mass 1 kilogram is equal to 10^{17} joules, which is just the amount of heat that would be released if we burned three million tons of coal. As we will soon learn, physicists have been able to release a small portion of this energy by smashing heavy atomic nuclei or forcing light nuclei to coalesce.

It must be stressed that the Einstein equation, $E = mc^2$, has to do not only with the energy inside nuclei. The equation is universal. But this is just like the spacemen's clocks. For the most part, the relationship between energy and mass cannot be verified. Indeed, take a ton of molybdenum and heat it to 1000 degrees. The mass will increase by 3 millionths of a gram. Only the fantastic magnitudes of intranuclear forces enable us to grasp the correctness of the equation $E=mc^2$.

This is a good time to warn the reader about sloppy statements of this remarkable equation. Some say that mass turns into energy; or, still worse, that matter is converted into energy. Actually, the formula $E=mc^2$ states that whatever mutual transformations of different types of matter occur, the change in energy that takes place in the system corresponds to an equivalent change in mass. Energy and mass are two uniquely related characteristics of matter.

Particles with Velocities Close to the Velocity of Light

The desire to reach down to the elementary building blocks of matter that make up the world is as old as the world. But for many long centuries this subject was at the mercy of the scholastic reasoning of wisemen. As soon as the opportunity appeared of actually breaking up molecules, atoms, and atomic nuclei, physicists took up the challenge with inspiration and persistence. The work goes on today and, to be quite frank, there appears to be no end in sight.

Obviously, to find out what the world is made of, one has to break up the component particles. That requires "projectiles", and the more energy they have the greater the hope of penetrating that secret of nature.

The history of the generation of fast particles began in 1932 when Rutherford's coworkers constructed a device to produce protons, which were accelerated to energies up to 500 kiloelectron volts. Then came cyclotrons capable of racing protons to energies that were measured in megaelectron volts (mega means a million). Next came the synchrotron that accelerated protons to a thousand million electron volts, and the era of giga volts was here (giga means a thousand million). Today, machines are being designed that plan to reach energies of millions of millions of electron volts. At a conference of physicists that took place in 1975 at Serpukhov (USSR), which has one of the biggest machines of that type, construction of a circular machine 16 kilometres in diameter was suggested.

The reader of course is already asking the obvious questions of how these machines work, why they are circular, and what the idea is all about anyway.

Actually, any vacuum device with a high voltage applied to the terminals can act as an accelerator of particles. The kinetic energy of particle accelerated to high velocity is equal to

$$\frac{mv^2}{2} = eU$$

(we have been making so much use of this formula that the reader has most likely got it firmly in mind by this time, which is all to the better). Even X-ray and television tubes can be regarded as accelerators.

But that principle of acceleration does not yield high velocities. The term "accelerator" is applied to machines capable of accelerating particles to near-light velocities. To attain that end, the particle has to be accelerated in succession through a number of fields. It is easy to see that a linear accelerator is not very convenient because a good many centimetres of path length is needed for some miserable tens of thousands of electron volts. In that way, reaching 10 000 million electron volts would require a path length of ten or so kilometres.

No, such a brute force approach to the problem is not suitable. In 1934 the American physicist Ernest Orlando Lawrence (1901-1958) laid the foundation for the construction of modern circular accelerators, which he christened *cyclotrons*. A single machine combines the acceleration of a particle by an electric field and its return many times to the original interval (gap) with the aid of a magnetic field.

The Lawrence accelerator looked like a tin can cut in two along a diameter. A rapidly alternating voltage is applied to the two halves, and the charged particles are accelerated during the time intervals in their traverse of the gaps between the halves. Inside the "tin can" we make the particles move in circular orbits by imposing a magnetic field whose lines of force are perpendicular to the bottom. In that case, as we know, the charged particle describes a circle of radius

$$R = \frac{mv}{eH}$$

The period of revolution is equal to

$$T = \frac{2\pi m}{eH}$$

For the electric field at the gaps of the machine to be able to pick up the particles, the variable voltage must be timed so that it changes sign at the very instant the particle arrives at the gap between the halves.

The charges are generated in the centre of the machine (for instance, ionization of hydrogen generates protons). The first circle is of small radius. But each successive circle increases the radius since, by the formula we just gave, it is proportional to the velocity of the particle.

At first glance it might appear that by increasing the dimensions of the cyclotron, and with it the radius of the circular trajectory, we could impart any desirable energy to the particle. And when the desired energy is attained, all we have to do is direct the beam out of the circle by means of a deflecting plate. This would be an ideal setup if it weren't for the dependence of mass on velocity. Einstein's equation for mass that did not seem to have any practical value now becomes the basis for designing circular accelerators.

Since the mass of a particle increases with velocity, the orbital period does not remain constant, it increases. The particle begins to come late and arrives at the accelerating gap not when the voltage phase changes 180 degrees but later. As the velocity increases, there comes a time when the electric field not only ceases to pick up the particles but even begins to slow them down.

A cyclotron permits speeding protons to about 20 megaelectron volts. Actually that is not very much. As I have said before, physicists are asking for more powerful machines. Obviously new approaches to the problem are needed if high energies are to be attained.

And the form of the equation for the orbital period suggests the direction to take. The mass increases with increasing velocity. Then what we need to do is increase the intensity of the magnetic field apace in order to maintain the period. True, this is a simple solution only at first glance. Don't forget that the radius of the orbit increases with each circuit of the particle. So it is required that the synchronous increase in mass and magnetic field hold good for a particle making successive circuits of progressively increasing radii. If we can disentangle this interrelationship of quantities, it will be clear that, first of all, there are "suitable" particles for which the condition will be fulfilled for a certain specified rate of increase in intensity of the magnetic field. What is most important, there will occur a peculiar kind of phase stability (automatic phase stabilization). A particle whose energy is greater than the energy needed for its orbital radius will slow down due to the excessive increase in mass; and, contrariwise, a lack of energy will lead to acceleration.

Very simple calculations using the formulas for the radius and orbital period of a particle will convince the reader that that is precisely the situation (specify the rate of increase in the strength of the magnetic field, compute the particle flight path, and then draw the curve, and you will see for yourself how the principle of phase stability functions). Or you can take my word for it that in this way we can, in principle, increase the velocity of particles without bound. We will have to use the pulse method of acceleration however. The machine operates when the field is increasing and is idle when in reverse. But we will not dwell on this method here. It too is out of date. If we kept to that principle, modern accelerators would have steel magnets weighing millions of tons.

Modern circular accelerators, called *proton synchrotron*, accomplish particle acceleration in a single orbit, and so the whole central portion of the magnet is cut out, as it were. These machines also operate on the pulse method. Here, both the intensity of the magnetic field and the period of the electric field vary in synchronism. Suitable particles build up speed as they move in a strictly circular orbit. The less suitable particles will oscillate about the good orbit but will also acquire a certain velocity.

In principle, fantastic accelerations can be achieved. For instance, proton velocities just short of the velocity light have been attained.

Now comes the question of why such machines are needed. Particle accelerators are constructed to learn more about the physics of elementary particles. The higher the energy of the charged particles used as projectiles to bombard targets the greater the chances are of discovering the laws of the mutual transformation of such elementary particles.

Roughly speaking, the world is made up of only three particles: electrons, protons, and neutrons. So far, the electron has no justification for its status as a component particle. As for protons and neutrons, it turns out that they can be broken up into pieces. New particles arise in diverse collisions that occur among the fragments. Today there are something like 250 such particles, and the great pity is that their number is constantly increasing (with the increasing power of accelerators). Specialists in the field of elementary-particle physics hope to develop a table of elementary particles like the Mendeleev table of elements and then reduce them to a small number of what might be called "protoparticles". After all, haven't the hundred and so chemical elements and several hundred isotopes been reduced to combinations of electrons, protons, and neutrons?

The reader then has every reason to ask: What meaning do we attribute to the phrase that the world is made up of three particles? The point is this. Only the proton and the electron are absolutely stable particles. The neutron is not quite stable, if the word "stable" is taken in its everyday usage. But its lifetime in the world of particles is tremendous: it is equal to approximately 10^3 seconds. As for the multitude of other elementary particles that are such a worry to theoreticians, their lifetimes come out to less than a millionth of a second. This is of course a very big difference.

Nevertheless we would like to bring these short-lived fragments of matter into some order. Many types of systems have been proposed for the elementary particles. But as soon as a more powerful accelerator comes into action and fresh phenomena are observed, the old schemes have to be discarded.

At the time of writing, I see that the specialists are optimistic. The entire system of elementary particles is being reduced to certain "protoparticles" called *quarks*. The only trouble is that quarks, unlike electrons and protons, have never been observed and, what is more, probably (in principle) never will be. In order to set up a "Mendeleev" table of elementary particles, the quark must be endowed with an electric charge either one third or two thirds of the electron charge, and it must have two additional parameters that are quite unimaginable. These parameters are called *strangeness* and *charm**.

I will not go into the problems that plague the elementary particles, and not because it is difficult to give a popular version of the existing schemes but simply because it is too early yet to be sure of their charm, so to say. It might very well be that entirely new ideas are in the making pertaining to elementary particles and quite new principles of approach to these minute portions (measuring in centimetres one over unity followed by fifteen zeros) of the universe.

Wave Mechanics

In 1923, in a scientific paper of exceptional boldness and with a streak of genius the French physicist Louis Victor de Broglie (1892-) wrote: "In optics, over the centuries, the corpuscular approach to phenomena has been far too neglected in comparison with the wave approach; might it not be that in the theory of microparticles the reverse mistake has been made?" De Broglie pointed out the path to follow in order to relate particles with wave conceptions.

His work was continued and expanded by the marvelous Austrian physicist Erwin Schrödinger (1887-1961). Somewhat later in 1926-1927, it became clear that the wave mechanics and quantum mechanics are actually equivalent terms. This new mechanics is one of the most important branches of physics that teaches us how to regard the behaviour of microparticles when neither their

---

*Since the manuscript of this book was handed in, new events have taken place in the world of elementary particles, a new parameter *colour* has appeared.

corpuscular aspect nor their wave aspect suffice to interpret events.

We warned the reader not to take too literally the expression "electromagnetic wave". Radio emission, light, and X rays can all be regarded in two aspects: wave and particle (corpuscular). The very same holds true for fluxes of particles. Although particle fluxes have clear-cut peculiarities that distinguish them from electromagnetic radiation (the main being that electrons, nuclei, neutrons, and ions can have a whole range of velocities, while photons can only move at a velocity of 300 000 km/s), this kind of matter also exhibits wave properties in certain situations and corpuscular (particle) properties in others.

What is the wavelength that must be attributed to a moving particle? By means of the following (somewhat simplified) reasoning, de Broglie demonstrated (rather, conjectured) what wavelengths are associated with what particle fluxes.

Let us examine the basic relations that connect the particle aspect of electromagnetic radiation with the wave aspect. A portion of energy of electromagnetic radiation carried by a photon is given by the formula $E=h\nu$. The energy of a photon, like that of any other portion of matter, obeys Einstein's equation. Thus, the photon energy can also be expressed by the equation $E=mc^2$. From this it follows that the mass of the photon is $m=h\nu/c^2$. Multiplying together the mass and the velocity, we obtain the value for the momentum of the photon as*

$$p=\frac{h\nu}{c}=\frac{h}{\lambda}$$

*The mass of the photon is the mass of a moving particle; now the rest mass of the photon is practically zero; experimenters claim it is less than 0.6×10^{-20} megaelectron volt. Also note that the relation for the momentum of the photon may be verified directly by measuring the pressure of light.

But we want to know the wavelength of a particle whose rest mass is different from zero. What can that be? Suppose the foregoing reasoning holds; assume that the relationship between the momentum and the wavelength is universal! It then remains to rewrite this expression in the form

$$\lambda = \frac{h}{mv}$$

This is the famous *de Broglie equation* (or *relation*). It shows that the wave aspect of a flux of particles must be revealed with particular clarity when the mass and the velocity of the particle are small. This is confirmed by experiment because it is easy to observe the diffraction of particles in the case of electrons and slow neutrons.

Verification of the truth of the foregoing, which incidentally in its day was regarded as a play on concepts, is quite straightforward. From the same substance, take an X-ray photograph (roentgenogram), an electron diffraction pattern, and a neutron diffraction pattern. If the particle velocities are matched so that the wavelengths in all cases are the same, then we should obtain identical (with regard to the radii of the rings) Debye crystallograms. And that is precisely what happens.

In 1927, an accident brought about the first verification of the de Broglie relation. Two American physicists, Clinton Joseph Davisson (1881-1958) and Lester Halbert Germer (1896-), were conducting experiments in the scattering of electrons on the surface of metals and, when working with their instrument, they happened to heat up the object. The object was a polycrystal; after being heated it was recrystallized, and so now the rays were scattered by a monocrystal. The pattern obtained was so much like appropriate roentgenograms that there was no doubt that electrons are capable of diffracting just like X rays.

Rather soon, observations of electron diffraction turned into a method of investigating the structure of substances, and in many instances it was more suitable than X-ray diffraction analysis. The main application of *electron diffraction analysis* is the study of the structure of thin films. The principles don't differ at all from the principles we discussed in Chapter 3. The difference lies in the fact that electron rays are scattered by electrons and nuclei, whereas X rays are scattered by electrons alone.

Since the wavelength of a particle is inversely proportional to the mass, it is clear that the diffraction of molecules is observed with difficulty. At any rate, it has not been achieved yet. It is possible to observe the diffraction of protons, but it is of no interest: protons do not suffice for studying volume structure due to the low penetrating power, and for surface studies it is better to use the diffraction of electrons because it yields far more information about the structure.

With neutrons it's different. Research into the diffraction of these particles has become a point of interest to hundreds of scientists. It is called *neutron diffraction analysis*.

It is technically much more difficult to obtain a neutron diffraction pattern than a roentgenogram. First of all, a sufficiently strong flux of neutrons of suitable wavelength (wavelength is regulated by neutron velocity) can only be generated by taking these particles out of a nuclear reactor through a special channel. The second difficulty is that neutrons are not easily scattered; they can readily pass through a substance without colliding with the atomic nuclei. That makes it necessary to work with large crystals of the order of a centimetre in size. It is not easy to obtain such crystals. Finally, the third difficulty is that neutrons do not leave traces in a photographic plate and reveal themselves in ionization chambers only indirectly. How neutrons are counted will be discussed later on.

Then why are scientists interested in neutron diffraction analysis? The point is that neutrons, unlike X rays, are not scattered by electrons but are deflected from their paths in encounters with atomic nuclei. There are many substances whose atoms differ only slightly as to number of electrons but differ radically as to properties of the nuclei. In such cases, X rays cannot distinguish the atoms but neutron diffraction analysis is highly successful. But perhaps the most important of all is that neutrons are strongly scattered by the nuclei of hydrogen atoms, whereas X rays can establish the positions of hydrogen atoms only with difficulty: the point is the hydrogen atom has only one electron. Now it is very important to know the position of this atom. In numerous organic and biological systems the hydrogen atom binds together the parts of a single molecule or adjacent molecules. This particular bond is called the hydrogen bond. Again, neutron diffraction analysis reigns supreme in distinguishing atomic nuclei having distinct magnetic properties. All these factors are sufficient to make neutron diffraction analysis an important method of studying the structure of substances.

The Heisenberg Uncertainty Principle

For a long time, many physicists could not reconcile themselves to the fact that light and particles simultaneously possess wave and corpuscular properties. This dualism seemed to contain something contradictory to the theory of knowledge. In particular, these scientists were opposed to the Heisenberg principle.

This most important proposition of physics of the microworld establishes the range of applicability of the corpuscular aspect of any phenomena connected with the motion of particles of a substance. The *Heisenberg uncertainty principle* can be written as follows:

$$\Delta x \; \Delta v > \frac{h}{m}$$

Here, Δx and Δv are the "smeared nature" of our knowledge of the coordinate and velocity of motion (in the direction of the same coordinate axis) of a blob of matter that we are considering in the corpuscular aspect. Briefly, Δx and Δv state the indeterminateness in our knowledge of the coordinate and the velocity of the particle.

It must be stressed at this point that we are not speaking of the technical difficulties of measuring. The foregoing relationship relates uncertainties that cannot be eliminated even in the most ideal experimental setup. Today, the various schemes proposed for an absolutely exact measurement of the trajectory and velocity of motion of particles are only of historical interest. A careful examination of them will always turn up a fundamental defect in the proposed scheme.

I will attempt, in a few words, to explain why experiment cannot yield greater accuracy than the Heisenberg principle allows for. Suppose we are discussing a definite position of a particle in space. To learn where it is located, the particle must be illuminated. As has already been mentioned, the possibility of distinguishing details is determined by the wavelength of the radiation used. The shorter the wavelength the better. But as we diminish the wavelength, so we increase the frequency of the light, and, hence, also the energy of the photon. The impact experienced by the particle under study will make it impossible for us to judge the velocity it had at the time of its encounter with the photon.

Or take this classical example. We place a narrow slit in the path of an electron. The electron passes through the slit and falls on a screen. A flash is seen. Thus, to within the accuracy of the width of the slit, we have established the position of the electron at the time it was passing through the aperture. Let us improve the precision. For this purpose we reduce the size of the slit. But then the wave properties of the electron will begin to take over (see page 57). The electron can deviate farther and farther away from the straight path, and that means that we will progressively lose more information about the component of its velocity in the direction of the plane in which the slit is made. Dozens of such cases can be thought up and we can consider them quantitatively (which is exactly what was done in the 1930s), and every time we arrive at the above formula.

Let us now discuss the estimates of Δx and Δv that can be made with respect to particles of different mass by using the Heisenberg inequality.

Suppose we are discussing an electron residing in an atom. Can we devise an experiment capable of establishing the site occupied by the electron at a given instant? Since the dimensions of an atom are of the order of 10^{-8} cm, we would like to obtain an accuracy of, say, 10^{-9} cm. In principle (true, only in principle) such an experiment is possible. But then let us estimate (using the inequality) the loss of information concerning the electron. For an electron, h/m is roughly equal to 7 cm²/s; for it, the Heisenberg uncertainty principle is written thus: $\Delta x\,\Delta v > 7$. And so Δv exceeds 7×10^9 cm/s, which is absolutely meaningless; in other words, we can't say anything about the velocity of the electron.

Well, and suppose we try to make a more exact estimate of the velocity of the atomic electron. For this purpose we can even think up an experiment that can actually be performed. But then we will have lost all knowledge about the site where the electron is located.

The inequality applied to an atomic electron shows that the corpuscular aspect does not operate in this case. The concept of the trajectory of an electron becomes meaningless and nothing at all can be said about the paths of transition of an electron from one energy level to another.

The situation changes when we seek the motion of an electron in ionization chambers. The track left by an electron is visible. So there is a trajectory after all. But then how is this linked up with the foregoing calculations? The answer is: there is no connection. One merely begins to reason anew. The track has a thickness of about 10^{-2} centimetre. This means the uncertainty with regard to velocity even for a slow electron going through a chamber at about 1 kilometre per second is quite acceptable. It is equal to 7 metres per second.

These numerical examples show us that the corpuscular aspect begins to vanish as soon as we attempt to look closer and inspect the portion of matter in more detail.

We can very often speak of protons and neutrons as particles, but if one is interested in their behaviour inside the atomic nucleus, which is about 10^{-13} centimetre in size, then the corpuscular aspect is not revealed.

And it is not hard to image that we can easily visualize the behaviour of a large molecule with molecular mass of the order of a million as if it were a pea. Such a molecule behaves like a good and honest particle. It is even possible to trace the trajectory of its thermal random motion.

The time has long since past when wave-particle dualism was regarded as something strange that requires a profound interpretation. Such outstanding scientists as Einstein and Bohr argued vociferously about how to interpret the "strange" behaviour of electrons and other particles. Today, the overwhelming majority of natural scientists see nothing particular in utilizing the two aspects in their descriptions of phenomena in which electrons, nuclei or photons participate.

Some ten years ago, a group of experts studying the work of scientists sent a questionnaire to a large group of physicists (about ten thousand). One of the questions was: Is the problem of the two aspects of matter of topical interest and has it been completely explored? Only twenty replied that they presume the Heisenberg inequality and its associated problems do not constitute the ultimate truth.

The difficulty in coming to terms with this important law of nature was probably due to a logical error underlying a protest that might be stated in the following words: "I cannot agree that the behaviour of particles of matter is unpredictable." The fallacy of this sentence lies in the fact that this parcel of matter is spoken of in the ordinary everyday sense of the word. Actually, the parcel of matter (we are speaking of light, microwaves, electrons or nuclei) is in no way like a grain. It is impossible to visualize a particle of matter; surely everyone will argee. Suffice it to say that to the electron or proton we cannot attribute colour, hardness, temperature ... . All these properties have to do with macroscopic bodies. Now if it is not possible to visualize a parcel of matter of this kind, all the more is it impossible to conceive of its motion. The motion of such a parcel of matter combines two aspects, the wave aspect and the corpuscular aspect. Therefore, only one of these aspects is unpredictable.

Quantum mechanics (and, to repeat, wave mechanics is a synonym) gives us a set of rigorous rules by means of which we can predict the behaviour of parcels of matter. A description of particles by the methods of quantum mechanics gives an exhaustive representation of the regularities of the microworld. With its aid we unerringly predict events and compel them to serve us practically.

This of course does not mean that new and more general laws of nature will not be discovered at some future time and that quantum mechanics may then become a particular instance of such generalities, just as Newtonian mechanics became a particular case of more general mechanics. Such general laws must be suitable for describing the behaviour of particles of small mass moving at high velocities. We are eagerly awaiting (and have been for quite some time) the creation of a theory that would unite all the "mechanics" into a single whole. There is already a name for this as yet "uncreated" theory; it is called *relativistic quantum mechanics.*

Is it not remarkable that the avalanche of discoveries made in the first quarter of the twentieth century has suddenly come to a halt? The reader may find it strange. But the fact remains. Despite the fantastic progress of applied sciences and despite the fact that the next two quarter-centuries have passed in the form of a scientific and technological revolution—despite all this, no new laws of nature have been opened up after the discovery of quantum mechanics. We will simply have to wait.

5. The Structure of Atomic Nuclei

Isotopes

In the third book of this series we told the story of how a flux of particles with different charge-to-mass ratios can be separated by means of electric and magnetic fields. And if the charges are the same, then separation of particles is possible as to mass. This is done by a device called a *mass spectrograph*, which is widely used for chemical analysis.

A diagram of this instrument is shown in Figure 5.1. The underlying idea is as follows. Particles with different velocities enter the electric field of a capacitor. Imagine a group of particles with the same e/m ratio. A flux of these particles enters an electric field and is separated into faster particles that exhibit a smaller deviation in the electric field and slower particles with a greater deviation. In fan-like fashion, these particles then enter a magnetic field perpendicular to the drawing. It is connected so as to deflect the particles in the opposite direction. Here too, the faster particles are deflected less than the slower particles. From this it follows that at some point outside the field, the imagined flux of identical particles will again collect in a single point, that is to say, they will come to a focus.

Particles with a different e/m value will also collect in a point, but that will be a different point. Calculations show that the foci for all e/m values will lie very close to a certain straight line. Now if a photographic plate

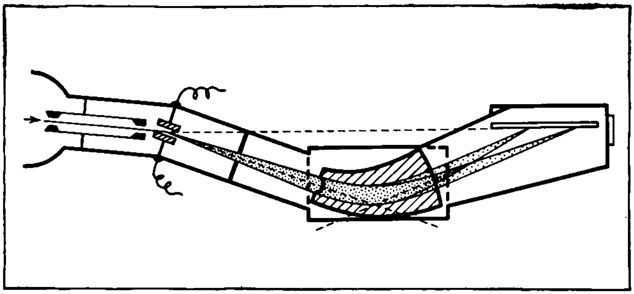


Figure 5.1

is positioned along this straight line, particles of each kind will reveal themselves by a separate line.

Isotopes were discovered with the help of a mass spectrograph. The honour for the discovery of isotopes goes to Sir Joseph John Thomson. In 1913, while studying the deflection of a beam of neon ions in an electric field and a magnetic field, he noticed that the beam was separated into two parts. The atomic weight of neon was known with sufficient accuracy: 20.200. It was found that in reality there are three kinds of neon atoms, with atomic weights 20, 21, and 22. Sorry, I'm rather used to the old terms; these numbers are now called relative atomic masses.

Since the chemical properties of neon do not depend on their masses, physicists were soon confident that the differences are connected only with the nucleus. The charge of the nucleus and the number of electrons are the same and so different kinds of atoms of neon should occupy the same place in Mendeleev's periodic table of elements. Whence the name, *isotopes*, or those that occupy the same place.

In the 1920s, the mass spectrograph took on its present-day features and a study began of the isotopic composition of all elements. All elements, without exception, constitute a mixture of isotopes. There are some, like hydrogen and oxygen, that consist mainly of one isotope (hydrogen of mass 1—99.986%, oxygen of mass 16—99.76%). There are others that contain different isotopes in about equal quantities. Say, chlorine (75% is an isotope of mass 35 and 25% is an isotope of mass 37). There are still other elements that consist of a large number of isotopes. The examples we have given are those of stable isotopes. Radioactive isotopes of an element (these are not stable and decay) will be discussed later on.

The mass spectrograph was soon refined to the point where it was established that the masses of isotopes are expressed by whole numbers only up to the second to fourth decimal place. The reasons for this deviation will be taken up later on.

The mass of an atom rounded off to a whole number is called the *mass number*.

Since the nuclear mass does not affect the chemical behaviour of an element, there are obviously many chemical compounds that differ in their isotopic composition. For example, there are two kinds of water, ordinary water and heavy water. In ordinary water, we find an isotope of hydrogen with mass number 1, in heavy water (called deuterium), the isotope of hydrogen has mass number 2. In the natural state, we find three isotopes of oxygen (mass numbers 16, 17, and 18), which means that water is a mixture of molecules of six different kinds. If the molecules of a substance consist of a large number of different atoms, then the number of isotopic varieties may run into the tens and hundreds.

Separating isotopes is an important branch of industry. And it is particularly important in a number of processes involved in the generation of atomic energy. Heavy water has to be separated from ordinary (light) water, different types of atoms of the nuclear fuel (uranium and thorium) have to be separated as well. And there are many many more such industrial problems that confront the physicist.

The complexity of the problem is that the atoms exhibit extremely small differences in their electronic makeup and, hence, in their chemical properties. In the case of light atoms, the separation procedure is carried out with extreme difficulty in the form of a multistage chemical extraction process. In the case of heavy atoms, the only possible way is to apply physical methods that make use of slight differences in the masses of atomic nuclei.

The most widely employed method today is the gaseous diffusion method. Molecules containing isotopes of different mass differ slightly as to their rates of passage through a porous barrier. Light-weight molecules get through faster than the heavier varieties.

Of course, the separation procedure could be based on the principle of the mass spectrograph that we have just discussed. But both methods are time- and money-consuming.

Only just two or three years ago it was demonstrated that isotope separation can be handled by a new laser method. The most important advantage here is that the laser generates a ray of extremely high monochromicity. Now the difference in distances between energy levels occupied by electrons in two isotopic varieties of the same element is naturally very slight. This is due to the mass difference of the nuclei since the charges on the nuclei of two isotopes are the same. And it is precisely the charges that determine, on the whole, the location of the electronic levels. Now a laser beam is so strictly monochromatic that it is capable of exciting only one kind of isotope while leaving atoms of the other variety unexcited.

Figure 5.2 depicts two processes for separating isotopes

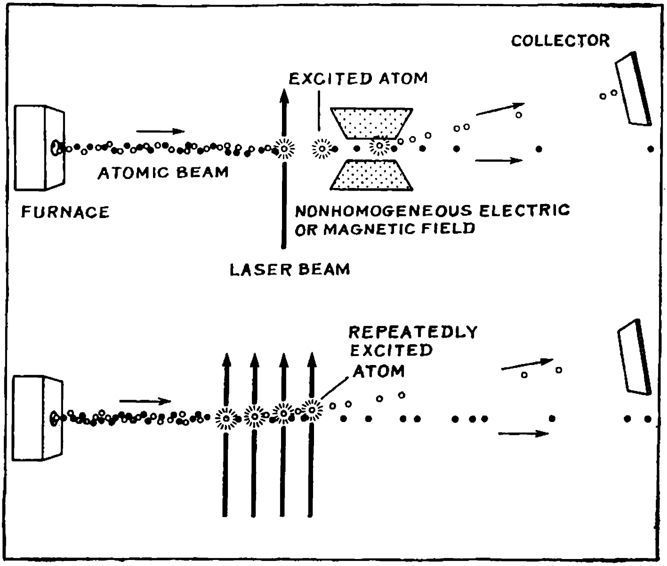


Figure 5.2

by means of a laser. A gas made up of atoms or molecules emerge from an opening in a furnace. The laser beam excites the atoms of one isotopic variety. As a rule, the excited atoms possess an electric or magnetic moment, and so a nonhomogeneous magnetic or electric field will deflect them to one side (upper diagram).

A second method is used if the excited atoms are rapidly de-excited. In this case, the same atom passes through the space covered by a laser beam and is excited a second time, thus several times experiencing inelastic collisions with photons. Each absorption of a photon imparts to the atom a momentum that is in the direction of the action of the laser beam. Atoms that can be excited are simply pushed upwards whereas atoms of the variety that does not absorb photons are not deflected in their motion.

The first successful experiment of this kind was carried out with a beam of barium atoms irradiated with a laser beam of wavelength 0.55535 micrometre. Absorption of a single photon shifted the atom through 0.8 centimetre per second in the case of a longitudinal velocity of 50 000 cm/s.

Radioactivity

In the third book of this series a short description was given of how Rutherford established that the atom consists of a minute nucleus and of electrons moving around the nucleus. Now comes one of the most important chapters in physics. It describes the structure of the atomic nucleus, which is made up of protons and neutrons. Strange as it may seem, the history of this discovery began fifteen years before Rutherford demonstrated his nuclear model of the atom in experiments devoted to the scattering of alpha particles by a thin piece of foil.

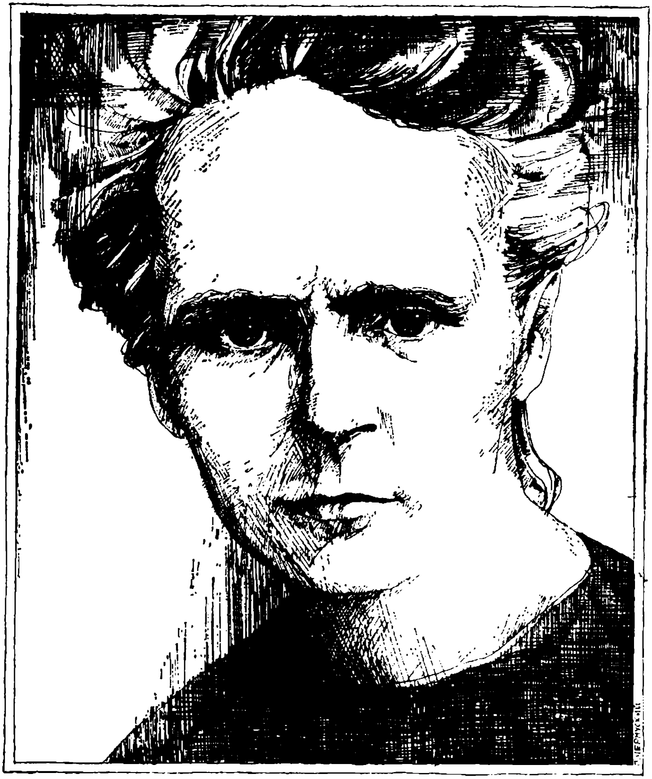
In the spring of 1896, the French physicist Antoine Henri Becquerel (1852-1908) found that uranium emits rays that act much like X rays. Just like the X rays of Roentgen discovered several months before, the uranium rays fog photographic plates and pass through opaque objects. Their absorption is proportional to the density of the object placed between the uranium and a photographic plate. If the body is opaque to these rays, the outlines of the object are formed on the plate. The uranium rays, like X rays, are capable of ionizing air; the amount of ionization of the air is a good measure of their intensity.

What is similar in the discoveries of Becquerel and Roentgen is the element of chance: they were accidental.

But an accident all by itself is never the source of an important scientific discovery. Just as there were people who had "seen" X rays several years before Roentgen, so there were (as it turned out following Becquerel's discovery) at least three people who had noticed the blackening of a photographic plate that had happened to be near uranium salts. But to "see" is one thing, and to pay attention to and locate the actual cause of the phenomenon is quite another thing. That is precisely what Roentgen and Becquerel did and what their predecessors failed to do. Hence the honour and the glory go to them.

The path to Becquerel's discovery went through the following stages. In the first tubes, the X rays fell on glass. The glass fluoresced under the action of cathode rays. And so it was natural to think that the penetrating rays accompanied fluorescence. Becquerel began by conducting experiments with substances that exhibit fluorescence under the action of the sun's rays. He was rather quick to see that the penetrating rays emanated from a variety of minerals containing uranium. That was a discovery. But Becquerel was in no hurry to report what he had found to the scientific community. The experiments were repeated a number of times. And it happened that the sun, out of spite, refused to come out of the clouds for several days. The photographic plates together with the mineral specimens lay in a drawer of his laboratory desk for a few days. On March 1, 1896, the sun finally came out and he could resume his experiments. But before doing so, Becquerel decided to check the quality of his plates. He went into the dark room, developed one of the plates and saw the clear-cut outlines of his mineral specimens. Now there had not been any fluorescence and so that could not have been the cause.

Becquerel repeated the "dark" experiments and was convinced that his minerals were the source of the pene-



Marie Sklodowska Curie (1867-1934)—outstanding woman scientist. In 1898, while studying the emission of uranium and thorium (the nature of the emission was not known at that time), she discovered in the ores of these elements certain substances that possess a far greater emission capability. She then isolated polonium and radium. Marie Curie and her husband Pierre Curie (1859-1906) introduced the term "radioactive". The discoveries of Marie Curie were immediately taken up by Rutherford and led to the laws of the radioactive decay of atoms.

trating radiation that develops all by itself, without any help from light outside.

A careful examination of many samples suggested to Becquerel that the source of the rays is uranium. If a mineral did not contain uranium, there was no penetrating radiation. And to complete the proof he decided to make a study of pure uranium. But uranium is a rare element. Becquerel asked his friend, the French chemist Henri Moissan (1852-1907), for some. At a meeting of the French Academy of Sciences, Moissan told of a method for obtaining pure uranium, and Becquerel reported that uranium emits rays. These papers were delivered on November 23, 1896. Only fifty years separate that discovery from the atomic bomb that was dropped on Hiroshima.

A year passed and in the autumn of 1897, two young physicists, the Curies—a man and wife team—began their experiments in a cold shed. But they worked with enthusiasm. Marie Sklodowska Curie (1867-1934) chose for the topic of her dissertation a study of the chemical peculiarities of mineral specimens that generate the penetrating radiation of Becquerel.

The intense work led from one discovery to another. First of all, it was found that, in addition to uranium, thorium also exhibits these penetrating rays. The intensity of the emission was measured by the intensity of the ionization current. Curie substantiated Becquerel's conjecture that the intensity of the penetrating rays does not depend on the composition of the chemical compounds involving uranium and thorium, and is strictly proportional to the number of their atoms.

Then came a setback: the uranium pitchblende ore yields four times more ionization than it should if it contained uranium. It is just at moments like these that the talent of the investigator is so important. An ordinary person would assume that the atoms of uranium were to blame. But Marie Curie realized that this phenomenon could be explained in a different way. It might be that the pitchblende ore contains a small quantity of some hitherto unknown chemical element capable of extremely powerful penetrating radiation.

The conjecture turned out to be true. The titanic work that Marie Curie carried out revealed a new element called *polonium* (in honour of Marie's home country Poland) and then *radium* (meaning ray). Radium turned out to be nearly a thousand times more active than pure uranium.

Now let us go more quickly, paying less attention to the historical sequence of events.

Following the discovery of radium, other substances were found that were sources of penetrating radiation. They all received the name "radioactive"

What is this radioactive radiation?

A radioactive specimen was placed in an evacuated box and this was surrounded by a lead shield with a slit. The rays passed through the slit, fell on a photographic plate, and left a trace on it. But as soon as the box was placed between the poles of a magnet, the developed plate revealed three marks. The radioactive beam had split up into three rays. One ray was deflected as if it were a flux of negatively charged particles, a second ray constituted a flux of positively charged particles, and one ray was not deflected at all. It was apparently somehow related to the X rays.

By methods we have already discussed it was demonstrated that, in the general case, radioactive radiation consists of a flux of electrons (before it was known that they are electrons they were called *beta rays*), a flux of atomic nuclei of helium (called *alpha particles*), and some kind of hard electromagnetic radiation (called *gamma rays*).

Radioactive Decay

Does something happen to the atoms that are sources of radioactive radiation? Yes, indeed. And these events are quite amazing. In 1902, our old acquaintance Sir Ernest Rutherford (1871-1937) (we have already told—disregarding the historical sequence of events—about his discovery of the structure of the atom in 1911) demonstrated that as a result of radioactive radiation there occurs a transformation of one kind of atom into another kind.

Rutherford expected that this supposition, though based on rigorous experimental proof, would be challenged vigorously by chemists. True enough. The point is that by proving the transformation of atoms, we encroach on the holy of holies—the indivisibility of the atom. By asserting that we can obtain lead from uranium we are accomplishing the dream of alchemists, who have merited as much "glory" as astrologists.

But under the weight of proof, the critics quickly retreated, and after some time the phenomenon of natural radioactive decay of certain atoms was incontestably demonstrated both by chemical and physical methods. What is the essence of radioactive transformation?

To begin with, it turned out that the electron rays that make up part of the radioactive radiation emanate from the nucleus. But if that is so, the charge on the nucleus increases by unity and the radioactive atom is converted into an atom next in order in the periodic table of elements.

An alpha particle carries a double positive charge and has a mass four times that of the hydrogen atom. If a nucleus ejects such particles, the atom must be "displaced" to the left in the periodic table of elements with an accompanying isotopic transformation.

It is almost too trivial to say that unstable atoms are subject to radioactive disintegration.

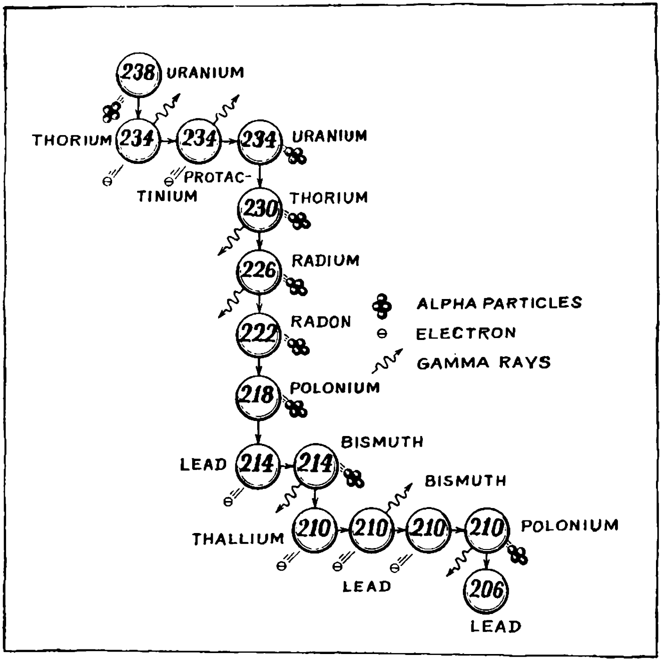


Figure 5.3

We do not know whether there were many such types of atoms when the earth began to cool off. But we have an excellent idea of what types of unstable atoms can now be found in nature. It turns out that they are members of three families, the progenitors of which are the uranium atom of mass number 238, the uranium atom of mass number 235, and the thorium atom of mass number 232.

Figure 5.3 depicts the first family. The first transformation is the transition of ^{238}U to ^{234}Th; this occurs due to the ejection of alpha particles. This is followed by two beta transformations that carry thorium to protactinium and then protactinium to uranium, but this time it is a uranium isotope of mass number 234. Then follow five consecutive alpha transformations that carry us down to the unstable isotope of lead of mass number 214. Another two zigzags and the process of disintegration comes to a halt: the lead isotope of mass number 206 is stable.

The destruction of each single atom is a random event. There are atoms that are "lucky" and have a long lifetime, others live for a fleeting moment.

But it is impossible to guess in any specific case when the given atom will be converted. We cannot name the date of demise of our cat or dog, but every animal species has its mean life span. And the same goes for every kind of atom: each has a very strict mean time of existence. Of course, atomic behaviour differs fundamentally from animal life. The life of unstable atoms, in contrast to the mean life span of living beings, does not depend on any external conditions. Nothing at all can affect what is called the mean decay time. Always, the same portion of atoms decays in unit time:

$$\frac{\Delta N}{N} = \lambda t$$

This formula holds true only for the case where the fraction $\Delta N/N$ is small.

The quantity λ is a constant for every radioactive transition. Instead of employing that constant, it is more pictorial to characterize the rate of the process by the "half-life", which is the time during which half of any given quantity of radioactive material is transformed. For different radioactive elements, this time can have an enormous range. For example, the half-life of the progenitor of the ^{238}U family is equal to 4.5 thousand million years. Contrast that with the fact that half of

Ernest Rutherford (1871-1937)—the eminent British physicist and a great experimenter. In delicate and highly original experiments he demonstrated the essence of radioactive decay. In his classical experiments in the scattering of a flux of alpha particles by a substance, he substantiated the modern theory of the structure of the atom as a system consisting of a nucleus and electrons moving about the nucleus. Continuing his experiments involving the bombardment of different targets with nuclei, he was the first to achieve an artificial transmutation of elements.

the atoms of the lead isotope of mass number 214 decays in one millionth of a second.

Nuclear Reactions and the Discovery of the Neutron

Radioactive transformations are quite similar to the chemical reaction of disintegration. Take a chemical substance, subject it to heat or light, and we obtain two substances. Say, carbonic acid breaks up into water and carbon dioxide, which is like what we have just considered: a thorium nucleus of mass number 230 decays into a radium nucleus and a helium nucleus.

If nuclear disintegration is possible, then there must be nuclear reactions that take place via the principle

$$A+B \rightarrow C+D$$

To obtain a chemical reaction of this kind, we have to mix the molecules of substances A and B. To achieve a nuclear reaction, we have to bring into collision two atomic nuclei.

Those were the experiments that Ernest Rutherford (1871-1937) began in 1919. This was before the time of particle accelerators, and nuclear reactions were achieved by means of bombarding a substance with alpha particles. After powerful fluxes of protons and other nuclei were obtained, new nuclear reactions were discovered. It became clear that in principle an isotope of any chemical element could be converted into another isotope. Even gold could be obtained from other substances. The dream of alchemists had become reality. The first nuclear reaction discovered was of the type $A+B \rightarrow C+D$, it was the conversion of nitrogen and helium into oxygen and hydrogen. We can write down this reaction as follows:

$${}^{14}_{7}N + {}^{4}_{2}He \rightarrow {}^{17}_{8}O + {}^{1}_{1}H$$

Note that the sums of the superscripts and the sums of the subscripts remain unchanged. The subscripts indicate the charge on the nucleus, the superscripts the mass rounded off to a whole number, that is, the mass numbers. Thus, the law of conservation of electric charge is maintained rigorously. As we shall see later on, the law of conservation of mass is only approximate. And, finally, the sum of the mass numbers is preserved just as strictly as the charge.

As early as 1920, Rutherford suspected that there must be a particle devoid of electric charge and close to the proton in mass. Otherwise, he said, it is difficult to understand how a positively charged alpha particle can penetrate into a positively charged nucleus, since like charged particles repel.

The particle without a charge—the *neutron*—was discovered in 1932. It is easy to see why its discovery was "held up" The point is we see charged particles through the tracks they leave in a gas or a photographic emulsion due to ionization of the molecules they encounter in their path. But an electrically neutral particle does not interact with electrons and therefore it does not leave any tracks. We can judge the existence of neutrons only on the basis of secondary effects.

The neutron was discovered in the bombardment of beryllium with alpha particles. This reaction can be written down thus:

$$^{9}_{4}\mathrm{Be} + ^{4}_{2}\alpha \rightarrow ^{12}_{6}\mathrm{C} + ^{1}_{0}n$$

The symbol n stands for neutron. But how can we be sure of the existence of a particle that does not itself leave any traces? On the basis of its actions. Imagine an invisible billiard ball on a billiard table. A visible ball is rolling over the green cloth and suddenly, inexplicably, bounds off to the side. The physicist cannot suspect the laws of conservation of energy and momentum to have gone wrong. And so he concludes that the visible ball collided with an invisible ball. What is more, by applying the conservation laws he is able to determine all the characteristics of the invisible ball; this he does by computing the angle of deviation from the line of flight and the change in velocity of the visible ball.

The number of neutrons is calculated in the following way. A substance containing boron atoms is placed in the path of a neutron beam. When a neutron encounters a boron nucleus it ceases to exist. The reaction that occurs is the following:

$${}^{10}_{5}B + {}^{1}_{0}n \rightarrow {}^{7}_{3}Li + {}^{4}_{2}\alpha$$

The neutron vanished but an alpha particle came into being. By recording these charged particles that leave visible tracks in a variety of devices, we can make exact measurements of the intensity of a neutron beam.

There are many other methods for determining, with complete assurance, all the parameters that characterize a neutron and, generally, any electrically neutral particle. An assemblage of precisely matched indirect findings is sometimes no less convincing than actually seeing visible traces.

Properties of Atomic Nuclei

Prior to the discovery of the neutron, physicists pictured the atomic nucleus as consisting of electrons and protons. This involved many contradictions, and attempts to create a theory of nuclear structure were all failures. As soon as the neutron was found in nuclear collisions, the idea immediately arose that the atomic nucleus consists of neutrons and protons. This hypothesis was first advanced by the Soviet physicist D. D. Ivanenko.

It was clear from the very start that the neutron mass was very close, if not equal, to the proton mass. This immediately gave rise to a clear-cut interpretation of the differences between isotopes of one and the same element.

As we see, to each isotope we can ascribe two numbers. One is the ordinal number in the periodic table of elements, Z, which is equal to the number of protons in the nucleus. The ordinal (or atomic) number determines the number of electrons associated with the nucleus. Now if that is so, then it is clear that the atomic number must be responsible for the chemical behaviour of the elements (this is because chemical reactions do not involve nuclei).

Now the mass number is equal to the total number of neutrons and protons, and so isotopes of one and the same element differ as to the number of neutrons in the nucleus.

Very precise experiments have elicited the characteristics of both particles that make up atomic nuclei. The mass of the proton is equal to 1.6726×10^{-24} gram, which is 1836 times more massive than the electron. The proton has a spin of 1/2 and a magnetic moment equal to 1.41×10^{-23} unit in the centimetre-gram-second system. The mass of the neutron is slightly greater than the mass of the proton, namely, 1.6749×10^{-24} gram. The neutron has spin 1/2. The magnetic moment of the neutron is antiparallel to the spin and is equal to 0.966×10^{-23} unit in the centimetre-gram-second system.

The spins and magnetic moments of atomic nuclei are studied by a variety of methods: use is made of optical spectroscopy, radiospectroscopy, studies of the deflection of beams of particles in a nonhomogeneous magnetic field. The general principles of these methods were discussed in the third book of this series and in the preceding chapters of this book. Here we will confine ourselves merely to a presentation of the basic facts obtained in the past few decades by large band of physicists.

First of all, I would like to stress that the laws of quantum physics pertaining to the moment of momentum (or angular momentum) hold true for all particles. And so for atomic nuclei we can write the formula for the angular momentum as:

$$\sqrt{S\ (S+1)}\ \frac{h}{2\pi}$$

Here, the quantity h is the Planck constant that we encounter in all formulas of quantum physics.

Usually the spin is the parameter S and not this expression. Theory states rigorously and experiment demonstrates brilliantly that the spin of every particle has to be equal to 0, 1/2, 1, 3/2, and so forth.

Looking through the tables of values of spin of different atomic nuclei (obtained in a variety of experiments), we find a number of interesting regularities. First of all, in nuclei with an even number of protons and an even number of neutrons, the spin of the nucleus is equal to zero (He, ^{12}C, ^{16}O). The number of nucleons (that is, nuclear particles) that is a multiple of four apparently plays a very big role. In many cases (though not in all) the spin of the atomic nucleus may be obtained as follows: drop the number closest to the mass number A that is a multiple of four and multiply the difference by 1/2. For example: in lithium-6 the spin is $2\times 1/2=1$; it is 3/2 in lithium-7, 1 in boron-10, and 3/2 in boron-11.

The rule is rather obvious: nuclei with an even mass number A have integral spin or spin zero, nuclei with odd A have spin equal to a multiple of 1/2.

The *Pauli exclusion principle* is applicable to both protons and neutrons in the nucleus. Two identical particles can reside in one energy level only if the spins are antiparallel. Since the proton and the neutron are different particles, one level can accommodate two protons and two neutrons. In this compact group with spin zero we perceive the helium atom (the alpha particle).

The presence of spin means the presence of a magnetic moment. As we know, there is a relationship of direct proportionality between the mechanical momentum L and the magnetic moment M. Here, the magnetic moment may be either parallel or antiparallel to the spin.

Bosons and Fermions

We have emphasized time and again that one energy level can accommodate only two particles with opposite spins. The time has come to say that this principle (the Pauli exclusion principle) holds true only for one class of particles. They are called *fermions*. The fermions include electrons, protons, and neutrons, and also all other particles that consist of an odd number of fermions. There is a second class of particles called *bosons*. The bosons include the photon, a number of short-lived elementary particles (like, say, the pi-meson) and (most important of all) all particles that consist of an even number of fermions.

There is no limit to the number of bosons that can occupy one energy level. To get a better grasp of the difference between bosons and fermions, take a look at Figure 5.4. Here, each black dot symbolizes a pair of particles with opposite spins. At very low temperatures, bosons mainly congregate on the lowest possible energy level. In this drawing, fermions appear as a column.

It is quite obvious that the differences in the behaviour of fermions and bosons are most obvious at low temperatures. At the very lowest temperatures, the number of bosons located in the "cellar" may be almost equal to the total number of bosons.

Don't try to "understand" what we have just said. Just remember it, because what we have said is really the ultimate truth. But I am very sorry every time I have to tell the reader (without offering proof) about things

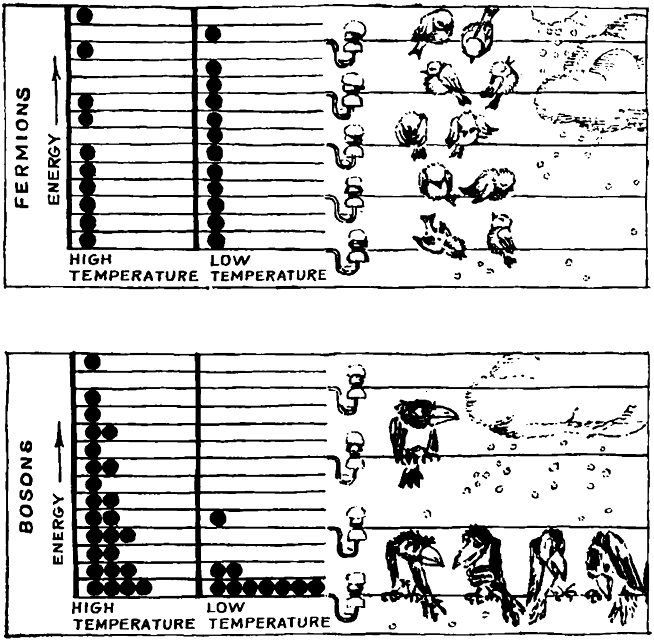


Figure 5.4

that can only be demonstrated with the aid of very involved mathematical equations. It turns out that in certain cases bosons can, in other cases cannot, congregate on a single energy level in large numbers. If they can, we say a *Bose-Einstein condensation* has taken place.

When a large number of particles congregate on one level, their motion becomes ideally matched. Twin particles move about identically despite the thermal chaos.

In the second book of this series we spoke of a marvelous liquid that possesses superfluidity at low temperatures. This liquid is a collection of ^{4}He (helium) atoms.

The atoms of this isotope are bosons. At a temperature of 2.19 degrees above absolute zero there occurs a condensation of the particles that imparts to the liquid an amazing property: superfluidity. The loss of friction may be explained in very rough fashion as follows: if only one atom gets through an extremely narrow slit, all the others follow suit immediately.

Actually we are dealing with two phenomena, where a flux of particles moves without regard for obstacles. The superfluid flow of ^{4}He atoms resembles electric current with zero resistance which is found in many metals and alloys and also at low temperatures.

But electrons are fermions. They cannot congregate together. The way out of this impasse was found in 1956 when three American scientists proposed a theory in accordance with which electrons can form into pairs below a certain temperature; and as we pointed out at the very start, a pair of fermions is a boson. Consequently, superconductivity sets in when such bosons condense on a single energy level. These two remarkable phenomena, *superconductivity* and *superfluidity*, have one and the same explanation. A single particle picks out a path that is "more suitable" and all other particles follow it.

If the idea of the conversion of fermions into bosons due to coupling into pairs is true, then we may ask: Can the isotope of helium of mass number 3, which has spin and is a fermion, become superfluid like ^{4}He?

It was evident from the very beginning that if this phenomenon does exist, then at any rate at temperatures much lower than the temperature of the transition of the basic isotope of helium, ^{4}He, into the superfluid state. The reason is clear: the nucleus of the ^{3}He atom consists of two protons and one neutron, which means it is 25% lighter than its brother. Therefore, naturally, its thermal motion will be more intense and setting up a march of bosons will become possible at lower temperatures. But

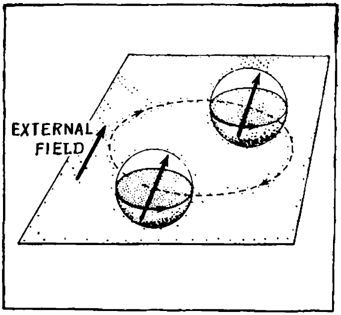


Figure 5.5

at what temperatures? Unfortunately, theory could not predict the transition temperature of ^{3}He to the superfluid state. Fantastic persistence and the overcoming of enormous difficulties were needed before superfluid (^{3}He) helium was obtained in 1974.

And at what temperature does the transition occur? It would be most appropriate to print the answer in boldface type: at 0.0027 degree Kelvin. And if the reader thinks that the two-degree difference between that and ordinary ^{4}He helium is nothing much to speak of, he has another thought coming. This is not at all like the difference between 20° C and 18° C. In this everyday event, the temperature fell by a factor of 293/291, whereas in the case we are discussing it dropped 1000 times—a tremendous achievement of experimental physics and a triumph of theoretical physics that predicted the coupling of atoms of ^{3}He into a "boson" pair.

A visual image might help in understanding and remembering this. Take a look at Figure 5.5. The magnetic moments of two atoms are in the same direction. Thus, the transition of ^{3}He into a state of Bose-Einstein condensation must be accompanied by a jump-like change in the frequency of the magnetic resonance. The point is that the pair behaves like a single whole, which is precisely what was detected in the experiment. This was such a brilliant page in the history of physics that it would have been a pity not to have told the reader about it despite any possibility of explaining under what conditions and on the basis of what events fermions couple up into boson pairs.

The Mass and Energy of an Atomic Nucleus

It was mentioned in passing that the mass number rounds the exact value of the mass of the nucleus to a whole number.

The accepted way today (this was discussed in the first book) is to choose the atomic mass unit as 1/12 of the mass of the carbon isotope ^{12}C.

The relative masses of the isotopes of all atoms differ from whole numbers very slightly but still so substantially that we cannot attribute these differences to experimental errors. The mass of ^{1}H is equal to 1.00807, the mass of deuterium is not at all twice as much, it is equal to 2.01463.

A careful study of the tables of isotopic masses shows that the mass of the nuclei is less than the sum of the masses of the elementary particles that constitute the new nucleus. For example, the mass of a neutron is 1.00888, the proton mass is 1.008807; the mass of two neutrons and two protons is equal to 4.0339, yet the mass of the nucleus of a helium atom, which consists of two neutrons and two protons, is not equal to that number, but to 4.0038. Thus the mass of the helium nucleus is less than the sum of the masses of the component particles of the nucleus by the amount of 0.0301 atomic unit, which is thousands of times greater than the measurement accuracy.

There can be no doubt that these small differences have a profound meaning. What is it?

The answer is given by the theory of relativity. Its appearance on the stage at that moment was undoubtedly more effective than when experiment confirmed the dependence of the electron mass on the velocity of the electron. The fact that the sum of the masses of the protons and neutrons making up the nucleus is less than the mass of the nucleus (this is termed the *mass defect*) is interpreted precisely and clearly with the aid of the famous equation $E=mc^2$. If a system acquires or loses an amount of energy ΔE, then the mass of that system increases (or, respectively, decreases) by the quantity

$$\Delta m = \frac{\Delta E}{c^2}$$

The mass defect of the nucleus (from the viewpoint of this principle) thus receives a very natural interpretation: it is the measure of the binding energy of the nuclear particles.

In chemistry and physics, the *binding energy* is understood to be the work which has to be expended in order to disrupt the given bond completely. If it were possible to split the nucleus up into elementary particles, then the mass of the system would increase by the amount of the mass defect Δm.

A breaking up of the nucleus would lead to the release of a tremendous amount of energy. It is simple to calculate that a change in mass by one thousandth of a relative unit, that is, by 1.66×10^{-27} gram, is equivalent to approximately 1 million electron volts.

Knowing the atomic mass of the nucleus, the reader will immediately see something very interesting. If we divide the energy binding the protons and neutrons in the nucleus by the number of particles, we get one and the same number, namely, 8 MeV (megaelectron volts, or million electron volts) for all nuclei, with the exception of some of the very lightest ones. From this there most definitely follows an important consequence: only the closest lying protons and neutrons interact, which means nuclear forces operate over short distances and become practically zero if one recedes from the proton or neutron to distances of the order of the dimensions of these particles (that is to say, 10^{-13} cm).

The quantity 8 MeV may be compared with the energies of chemical bonds of molecules. This comparison yields the interesting fact that the chemical bond energy is usually equal to several electron volts per atom. This means that several million times less energy is needed to break up a molecule into atoms than to disrupt (fission) the atomic nucleus.

From the foregoing it is clear that nuclear forces attain fantastic values. It is also obvious that nuclear forces constitute a new class of forces since they are capable of holding together particles that have like charges of electricity. Nuclear forces are reducible to electric forces.

The regularities that these two kinds of force obey are very highly disparate. Electromagnetic forces diminish slowly, and instruments record electromagnetic fields at enormous distances from charged particles. Nuclear forces, on the contrary, fall off very rapidly with distance. Practically speaking, they do not operate beyond the limits of the atomic nucleus.

Another important difference is that nuclear forces (very much like chemical valence forces) possess the property of saturation. Each nucleon (that is, proton or neutron) interacts with a limited number of its closest neighbours. There is no such limitation to the action of electromagnetic forces.

So there are three kinds of force in nature: gravitational, electromagnetic, and nuclear. Is that right? So far we do not know for certain. Physicists claim there is a fourth force with the rather inapt name of "weak interaction". We will not go into that matter here, all the more so since there is some hope that it will be reduced to the electromagnetic forces.

The Energy of Nuclear Reactions

We have illuminated two important facts. First, atomic nuclei can be involved in reactions that take place in accord with schemes that are very much like those familiar to chemists; second, the original nuclei and the newly generated particles will always differ slightly as to mass (this is because the sum of the mass numbers is preserved but not the sum of the masses of the nuclei prior to and after the reaction).

And besides we also saw that negligible differences in mass are accompanied by the release or absorption of tremendous amounts of energy:

The energies that are released or absorbed during nuclear transformations can in no way be compared with the heat released in chemical reactions. Let us take some examples to illustrate this point. One gram of coal is burned and releases heat sufficient to raise half a glass of water to the boiling point. Now the amount of heat generated in nuclear transformation is, as we said, no comparison: if it were possible to break up all the nuclei in one gram of beryllium using alpha particles, the heat released would suffice to bring to the boiling point one thousand tons of water.

This fact was well known to Rutherford and his co-workers but nevertheless Rutherford said he thought the utilization of nuclear reactions for practical purposes an impossibility (at that time, physicists had not the slightest inkling of the idea of chain reactions). We might point out here that he was just as incapable of foreseeing the coming revolution that his discovery had wrought as were Michael Faraday (1791-1867) and Heinrich Rudolf Hertz (1857-1894) concerning the interesting psychological conjecture mentioned in the third book of this series. But since we know what followed the modest experiments of Rutherford, we must take some time off to remind the reader of the mechanism of release and absorption of energy in reactions.

First, let me state the similarities between chemical and nuclear reactions.

Reactions of the type where particles A and B are converted into particles C and D release or absorb heat depending on whether slow particles are turned into fast particles or fast particles are turned into slow particles. That is the situation in chemical reactions and the same goes for nuclear reactions. Furthermore, if fast particles are created out of slow particles, this means the kinetic energy of the system increases. But the law of conservation of energy permits this only if the potential energy of the system has diminished. That is to say, in this case the sum of the internal energies of particles A and B is greater than the sum of the internal energies of particles C and D. That is the situation in chemical reactions and the same goes for the internal energies of nuclei.

According to the Einstein law, any reduction in the internal energy is uniquely related with a reduction in mass. An increase in the internal energy leads to an increase in mass. That is the situation in chemical reactions and the same goes for nuclear reactions.

But in chemistry the law of conservation of mass is operative. The sum of the masses of molecules A and B is equal to the sum of the masses of molecules C and D. Now, in nuclear reactions, that equality is not maintained. So there is a difference! No, there is not. The difference is only quantitative. In chemical transformations, the changes in energy (and hence in mass) are so slight (slight from the point of view of relativistic theory) that the changes in the masses of the molecules cannot be detected experimentally. Thus the analogy between both types of reactions is one hundred percent.

Because this is so important (many people think that the release of nuclear energy is some kind of special process, but that is not so), let us reason in similar fashion for the case where particle A decays into particles B and C. If the particle splits up into parts all by itself, then we say that particle A is unstable. If A is a molecule, we say that the substance decays (disintegrates). If A is a nucleus, we say the substance is radioactive. In both cases, heat is released. Particles B and C will possess some kind of kinetic energy that they did not possess before. This energy came from the potential energy. Pictorially speaking, the spring connecting particles B and C broke; to put it scientifically, the binding energy vanished. It was due to this binding energy that we obtained fast-moving particles B and C, that is, the energy was released in the form of heat.

Because of the small amount of energy, in a chemical reaction we do not detect any difference in the mass of molecule A and in the sum of the masses of molecules B and C that originate from A. Now in the case of nuclear reactions, this difference is readily detected experimentally. Nuclei B and C will have a mass exceeding the mass of nucleus A by the amount of the mass defect.

The very fact that a certain reaction generates heat does not yet mean that it will have some practical value. The condition of instability of a system, that is, the circumstance that the original substance is at a higher energy level than the products of the reaction, is, as the mathematician would say, a necessary condition but not a sufficient condition.

In book two we discussed at length the conditions that must be fulfilled for a substance to serve as a chemical fuel. Here we need merely continue the analogy between chemical and nuclear reactions.

To summarize, then: it is not enough that a chemical reaction yield heat, it is necessary that the heat "ignite" adjacent molecules.

It is therefore clear that having learned to make atomic nuclei collide with the release of tremendous amounts of energy, physicists had not in the least come close to the creation of nuclear fuel.

When alpha particles bombard beryllium or lithium, these elements do not behave like a fuel. They do, however, satisfy the first requirement of a fuel: they produce energy. Lithium and beryllium behave like tiny pieces of coal, each of which has to be ignited with a separate match.

Right up to the end of the 1930s, the creation of nuclear fuel was regarded as a totally hopeless undertaking.

A Nuclear Chain Reaction

Beginning from 1934, mainly due to the work of the Italian physicist Enrico Fermi (1901-1954) and his collaborators, evidence was obtained that the atomic nuclei of most elements are capable of absorbing slow neutrons and thus becoming radioactive.

At that time, certain radioactive transformations were known to consist in the emission of electrons and alpha particles (these transformations are accompanied by gamma radiation). But in 1938 a number of investigators (it is interesting to note that the fundamental discovery we are about to discuss was not done by one person) found that uranium activated by neutrons via Fermi's method contained an element similar to lanthanum. There could be only one explanation: under the action of neutrons, the uranium atom breaks up into two more or less equal parts. The exceptional importance of this discovery was

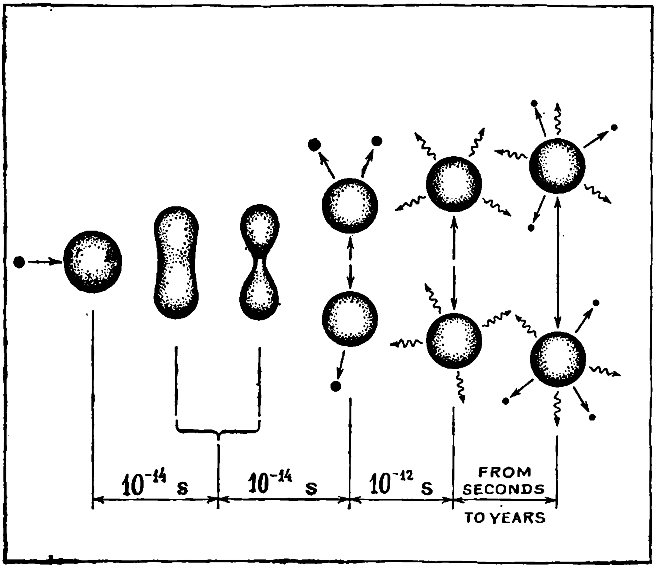


Figure 5.6

clear from the start. The point is that by that time the following regularity had been discovered: the greater the atomic number of an element the more neutrons there are in the nucleus. In uranium, the ratio of the number of neutrons to the number of protons is approximately equal to 1.6. And for elements like, say, lanthanum, located in the middle of the periodic table of elements, this ratio ranges between 1.2 and 1.4.

Now if the uranium nucleus fissions (that's the term) into two roughly equal halves, then the nuclei of the fission products will invariably contain some "extra" neutrons. They will eject neutrons, and neutrons play the role of "matches".

The possibility of a chain reaction was now evident.

The first calculations of this phenomenon were carried out in 1939. The dramatic sequence of events, from the first atomic pile (now called reactor) to the making of the atomic bomb and its explosion at Hiroshima, has been discussed in all its detail in dozens of books, but that is beyond the scope of our story and we will confine ourselves to the present-day state of the matter.

We have three questions to take up: first, the meaning of a nuclear chain reaction, second, how the reaction can be harnessed, and, third, when it results in an explosion.

Figure 5.6 is a diagram of one of the most important reactions of this type: the fission of the uranium-235 nucleus.

The question of the first neutron is simple: it can be found in the atmosphere. If we want a more active "match", we can take a tiny mixture of radium and beryllium.

A neutron enters the nucleus of a uranium-235 atom which consists of 92 protons and 143 neutrons packed tightly into a sphere of radius of about 10^{-12} cm and produces an isotope called uranium-236. The visitor deforms the nucleus and, after a time interval of about 10^{-14} second, the two halves of the nucleus are held together by a tiny bridge; in another 10^{-14} second the nucleus has split into two pieces. At the same time, each of the fragments ejects two or three (an average of 2.56) neutrons. The fragments fly apart with a colossal kinetic energy. One gram of uranium-235 produces as much energy as 2.5 tons of coal, or 22 000 kilowatt-hours. Then, 10^{-12} second later the nuclei formed in the fission process have more or less calmed down with the emission of eight photons of gamma radiation. The new nuclei are radioactive. Depending on the kind of fragments that have formed, the subsequent process of decay may continue for seconds or for many years with the emission of gamma rays and the ejection of electrons.

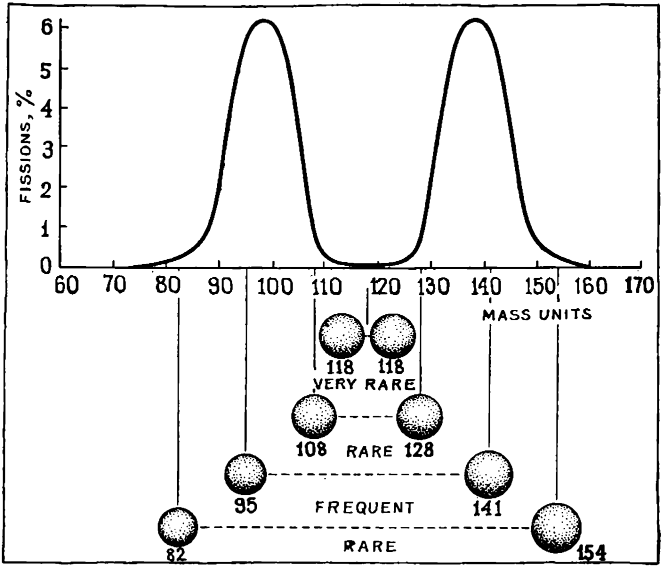


Figure 5.7

Figure 5.7 shows that for the most part the nucleus of uranium-235 splits into two unequal fragments. As is evident from a glance at the curve, most fissions result in the formation of nuclei with mass numbers 141 and 95.

The set of newly produced radioactive fragments is very great at any rate. The most varied requirements of industry with respect to artificial radioactive elements can now be satisfied.

If the neutrons formed in the fission of one nucleus can be made to fission the nuclei of other atoms of uranium, then a chain reaction can be set up.

Since matter is extremely "full of holes" as regards its nuclear structure, it is highly probable that the neutrons formed in the fission of a nucleus will leave the substance without fissioning any other nuclei. What is more, we must not forget that not every encounter between a neutron and a nucleus leads to fission. A chain reaction is ensured only if at each instant of time the number of neutrons located inside a piece of substance is the same or greater than the number at the preceding instant. This condition is stated by the physicist as follows: the neutron multiplication factor—which is equal to the product of the number of neutrons into the probability of a neutron encountering a nucleus and into the probability of neutron capture by the nucleus—must not be less than unity.

That is why pure atomic fuel has a critical mass. If this mass is less than critical, one can calmly (well, more or less calmly, let us say) carry a piece of nuclear fuel in one's pocket. And it won't be heavy because the critical mass is close to one kilogram.

Naturally, it is very important to know the figure for the critical mass. The first calculations of this quantity were carried out in 1939 by Francis Henri Perrin (1901-), the son of Jean Baptiste Perrin (1870-1942). This calculation is only of historical interest today because at that time it was not known that a chain reaction cannot develop in natural uranium, no matter how much of it is taken. But not much time was needed for the picture to become quite clear. A chain reaction does not develop in natural uranium because the neutrons produced in the fission of the nuclei of uranium-235 are absorbed via what is called "resonance" capture by the atoms of uranium-238 with the formation of uranium-239, which, by means of two successive beta disintegrations, turns into neptunium and plutonium. Only uranium-235 and plutonium possess a critical mass. Substances with a critical mass constitute nuclear fuel. These were the facts that physicists had at their disposal at the beginning of the 1940s.

If a device could be designed in which pressing a button brought together two pieces of nuclear fuel, each of which has a mass less than critical and, when joined, has a mass exceeding the critical mass, then an explosion would take place. That is the simple principle that underlies the working of the atomic bomb.

Now what do we have to do to control the course of a nuclear reaction? Quite obviously we have to create a system containing not only the atoms of the fuel but also other atoms capable of absorbing neutrons and, as it were, putting them out of action. Cadmium rods are quite suitable for this purpose. Cadmium is a good absorber of neutrons. If we make a structure consisting of rods of nuclear fuel and rods of cadmium and then arrange for inserting and extracting the rods (this takes place in the body of a nuclear reactor—at first it was called a pile), we can set up a nuclear chain reaction by making the neutron multiplication factor slightly more than unity; then we can raise the heat release to the desired level and insert cadmium rods so that the multiplication factor becomes exactly unity.

6. Energy Around Us

Sources of Energy

In the final analysis, it will be seen that all the energy generated here on the earth comes from the sun. Temperatures of the order of millions of degrees rage in the deep interior of the sun. That is where reactions develop between atomic nuclei. They go by the name uncontrolled thermonuclear fusion.

Man has harnessed these reactions and on the earth they release fabulous amounts of energy. The hydrogen bomb, for example. Research scientists are now attempting to bring the thermonuclear fusion reaction under control. That is the topic of this chapter.

Now that we have discussed the structure of nuclei we are ready to examine the sources of terrestrial energy.

Under terrestrial conditions, we can obtain energy in three ways. First, we can extract energy from fuel (chemical or nuclear). This requires setting up a chain reaction in which molecules or atomic nuclei are combined or disrupted. Chemical fuels of practical value include coal, petroleum, and natural gas. Nuclear fuel for fission (breaking up of particles) includes uranium and thorium, and nuclear fuel for fusion (combining of particles) constitutes primarily the light element of hydrogen.

Second, the conversion of kinetic energy into work. Rivers can be made to generate electricity and thus "work" Hydraulic power ("white coal") becomes one of the most important sources of energy. Here water falling from considerable heights generates electricity; this can be achieved by constructing dams or by utilizing natural waterfalls, like the Niagara Falls. Similarly, wind can be harnessed to convert energy into work turning the blades of a windmill. We shall see that wind ("blue coal") must be taken seriously. Windmills are coming to life again on a much higher engineering level and will soon be making a substantial contribution to the energy bank. The same can be said of such an energy source as tidal waves.

The laws of thermodynamics suggest a third solution to the energy problem. In principle, it is always possible to design an engine that can use temperature differences. As we know, a transfer of heat from a hot body to a cool body can be partially converted into mechanical work. Now we encounter temperature differences in many places: in the earth's crust, in the oceans, and in the atmo sphere. Going down into the interior of the earth (actually, of course, it amounts to digging slightly into the crust) we invariably observe an increase in temperature.

To repeat, all these three possibilities ultimately stem from the fact of solar radiation coming to the earth. The earth receives an insignificant portion of the energy radiated by solar rays. But even this minute bit is so colossal that for many years it will handle our needs and be enough to carry out the most fantastic projects.

But solar energy can also be utilized directly. The reader will recall that we have learned to convert radiant energy into electric current with the aid of photoelectric devices. As we know, the generation of electricity is the most important way of making energy serve practically.

Of course, in many cases both the internal energy of matter and the energy of the motion of water and wind can be utilized directly, bypassing the stage of conversion into electric current. However, perhaps in all cases, with the exception of aircraft and rockets, it is advisable to obtain electricity from the prime source at power plants and make the electric current serve our needs. The importance of electric energy will grow immeasurably as soon as we learn how to construct light-weight high-capacity storage batteries for electric automobiles in place of the heavy low-capacity ones now in use.

Before going on to a concrete discussion of energy sources, I wish to bring the attention of the reader once more to two important classifications of sources. First of all, there is an important dividing line between fuel and solar energy. In the former case, we are spending the wealth of the earth which cannot be replaced. Now the sun, the wind, and the water are sources for which we do not have to pay anything. They are free of charge in the sense that using their energies does not in any way diminish the wealth of the earth. Windmills do not reduce the amount of air in the atmosphere, the operation of hydropower plants does not take water out of the rivers, and solar-driven power-generating devices do not use up any terrestrial resources.

There is another classification. We have to protect nature at large—the flora and fauna of the earth. This is a problem whose importance is hard to overestimate. Burning fuel is bad both because it depletes the material of the earth and also because it pollutes the ground, water, and air with enormous quantities of deleterious waste. Even hydroelectric power plants have their drawbacks. Changes made in the water regimen of rivers can affect the climate and complicate the life cycles of the fish population of the world.

There can be no doubt that the optimal mode of obtaining energy is the direct utilization of the sun's radiation.

Such are the principles. Now let us take up a more concrete discussion of the possibilities of energy utilization under terrestrial conditions and give some of the figures that can be found in the reference literature on power generation.

We start with solar energy. Every square metre on the outer boundary of the earth's atmosphere receives about 1.4 kW of energy (which is about 10 million kilocalories of energy in one year). Hundreds of kilograms of coal would be needed to generate that much heat. So how much heat does the earth as a whole receive from the sun? Calculating the area of the earth and taking into account the nonuniform illumination of the earth's surface by the sun's rays, we obtain about 10^{14} kW. This is 100 000 times more energy than that produced by all the power sources on the earth: factories, power plants, automobile engines, aircraft engines; in other words, this is 100 000 times more power than the energy consumed by the entire population of the world (about a thousand million kilowatts).

Up to now, solar energy has hardly at all been touched. The point is this: although the overall figure is enormous, a good deal of the energy falls on the slopes of inaccessible mountains, on the oceans that occupy most of the earth's surface, and on the sands of desert areas. What is more, only a relatively small amount of energy falls on any small area. Surely it would not be feasible to construct energy receivers over many square kilometres of surface. And, finally, solar energy can be converted into heat only in localities where there are regularly many sunny days during the year.

The thirst for power and tremendous advances in the production of semiconductor photoelectric cells have radically changed the psychology of power engineers. There are many designs and even experimental units that focus solar radiation on thousands (in the future there will be millions, perhaps even thousands of millions) of photoelectric cells. And cloudy days and absorption of rays by the atmosphere are not serious deterrents. There can be no doubt that there is a great future in the direct utilization of solar energy.

Wind is now regarded in a different light too. Some twenty years ago wind was dismissed as a reasonable source of power. It has the same defect as solar energy, and that is that the amount of energy per unit area is relatively small. In order to generate power for factories, one would need windmills with vanes too large to be practicable.

Another essential drawback is that the wind force is not a constant factor. All that relegates the energy of the wind, or "blue coal" as the poet says, to small engines (windmills). When the wind is up, these power plants generate electricity for agricultural machines and home lighting. If extra power is generated, it can be stored in storage batteries, which will take up the load in times of calm. Obviously, the windmill has only a secondary role to play.

Today, engineers discussing the energy problem reason differently. Designs for electric power plants consisting of thousands of appropriately positioned "windmills" with enormous vanes are on the verge of being made into hardware. The use of wind will surely make an appreciable contribution to man's ever increasing needs.

Another gratuitous source of energy is moving water, the tidal waves of the ocean that constantly wash up onto the land, and the water of the rivers that flow to the seas and oceans. In 1969, hydroelectric power generation came to 115.2 thousand million kilowatt-hours in the USSR and 253.3 thousand million kilowatt-hours in the USA; however, utilization of water resources in the Soviet Union is only 10.5% while in the United States it comes to 37%.

These figures are impressive, but if we were deprived of all coal, petroleum, and other energy sources and only made use of hydro power—even if all engineering opportunities were utilized for hydroelectric power generation—we could not meet our requirements today.

What energy can be extracted from tidal waves? A great deal, although it is approximately one tenth the energy of rivers. Unfortunately, only a small portion of this energy can be harnessed. Because of the pulsating nature of the tides, there are many complications. However, Soviet and French engineers have found some practical ways of overcoming them. Today, tidal power stations ensure guaranteed outputs during hours of maximum consumption. In France, a tidal power station has been constructed on the river Rance, in the USSR there is a station at Kislaya Guba in the vicinity of Murmansk. This latter station is an experimental model for tidal power stations of 10 000 million watts (now in the design stage) to be built in bays of the White Sea.

Ocean water at considerable depths has temperatures 10° to 20° different from the surface layers. This offers possibilities of constructing a heat engine whose heater, in the medium latitudes, would be the upper layer of water and whose cooler would be the deep-lying layer. Such machines could have an efficiency of 1% to 2%. This of course is also a highly unconcentrated source of power.

Another gratuitous source of power is geothermal energy. We will not discuss countries that have many geysers. There are not so many; and where there are, the heat of the geysers is utilized for industrial purposes. What we shouldn't forget is that two or three kilometres below any spot on the globe we encounter temperatures of the order of 150 to 200 degrees Celsius. The principle underlying a geothermal power station is obvious. Drill two channels. Cold water enters one and hot water comes out of the other.

12—2542

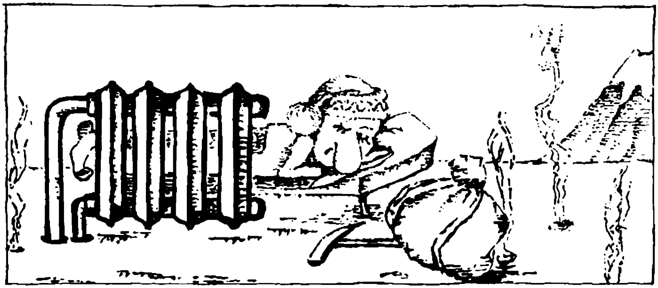


Figure 6.1

Fuel

All the power sources described above have great advantages over fuel. Fuel is burned. Utilizing the energy of coal, petroleum, and wood constitutes largely an irreplaceable destruction of terrestrial wealth.

What are the reserves of fuel here on earth? Ordinary fuel that burns when ignited with a match includes coal, petroleum, and natural gas. Their reserves are extremely limited. At the rate we are now using up petroleum, the known reserves will have been depleted completely by the beginning of the next century, the same goes for gas. There are somewhat larger reserves of coal, something like ten million million tons. The combustion of one kilogram of coal yields 7000 kilocalories of heat. (There are all kinds of fuel, and quality varies. This is naturally only a standard fuel used for measurement; it is a unit of comparison fuel used to compare different types of material.) Thus, the overall energy supplies of coal come to something on the order of 10^{20} kilocalories, which is roughly a thousand times the world's annual power consumption.

This thousand-year reserve is actually not much. A thousand years is a great deal if compared with the average human life span, but a person's life is only a moment when compared with the time of life on earth and with the existence of human civilization. What is more, power consumption is constantly increasing per head of the population. For this reason, if the fuel supplies consisted solely of petroleum and coal, the energy-reserve situation here on earth would have to be regarded as catastrophic.

But must we confine chemical fuel to the naturally found substances? Of course not. In a number of cases, it may prove that synthetic gaseous or liquid fuel is better than petroleum and natural gas.

In recent years much attention has been paid to the industrial production of hydrogen. As a fuel, hydrogen has many good points. It can be produced in limitless quantities in a variety of ways. It is readily available everywhere and so there are no transportation problems. Hydrogen can easily be purified of any undesirable impurities. In a number of cases it will be found that burning hydrogen directly for heat is the best thing. We can bypass the stage of converting it into electricity.

At present, there are three main profitable ways of obtaining hydrogen: electrolytic, thermochemical decomposition, and, finally, irradiation of hydrogen-containing compounds with neutrons, ultraviolet rays, etc. It is even economically profitable to obtain hydrogen from coal and petroleum in nuclear reactors. In these cases, the hydrogen can be delivered by pipeline to the site of consumption, as is done with natural gas.

Concluding this brief survey of chemical fuels, let us now ask the question: What is the situation like concerning nuclear fuel? What are the reserves of nuclear fuel on the earth? Remember that only small amounts are consumed in nuclear reactors. One kilogram of nuclear fuel produces 2.5 million times more power than a kilogram of coal.

Approximate estimates show that the reserves of potential nuclear fuel (why the word potential is used will be clear from what follows) come to about 2 million tons of uranium and 4 million tons of thorium. These are the fuels used to generate power in nuclear reactors based on the nuclear fission process. Will any other materials be added to these two elements? Perhaps. The number of nuclear reactions that produce energy is tremendous. The whole problem is how to make the process a chain reaction.

Let us first see what we are able to do at the present time. As follows from the preceding chapter, there is only one naturally found substance that can be used as nuclear fuel. That is the isotope uranium-235. The uranium mined in the earth contains 99.3% of uranium-238 and only 0.7% of uranium-235.

At first glance the simplest idea is this: isolate the necessary isotope, build reactors consisting of pieces (or rods) of that substance, introduce into the body of the reactor control rods that absorb neutrons and, thus, control the nuclear reaction.

First of all, note that by absorbing neutrons and not allowing them to participate in the chain reaction we reduce the power of the reactor; in other words, we extract smaller amounts of energy per second from a unit mass of nuclear fuel. Now, if we slow down neutrons to thermal velocities and convert the "fast" neutrons generated in the disintegration of a nucleus into "slow" neutrons, that will be extremely useful in enhancing the efficiency of the reactor. This is because the nuclei of uranium-235 absorb slow neutrons with a far greater probability.

In all designs (except experimental reactors that haven't gone beyond the laboratory stage), the moderator has been either heavy water or ordinary water. Heavy water is good because it does not absorb neutrons at all, but it is not so good at slowing down neutrons as ordinary water is.

Thus, the best and simplest method is to isolate the uranium-235 isotope. That, as you recall, is a very expensive procedure. Chemical methods are useless: after all, these are chemically identical substances.

At present the most profitable method is by centrifuging. Before this operation can be started, we have to obtain a gaseous compound of uranium. The only such compound that is in the gaseous state at ordinary temperatures is uranium hexafluoride. The mass difference of the gas molecules containing the isotopes uranium-238 and uranium-235 is so slight that the best centrifuge enriches the gas with lighter molecules only up to 12%. In order to obtain uranium with 3% of uranium-235 (such fuel can then be used in a nuclear reactor), the process has to be repeated 13 times. Clearly, the true solution of the engineering problem cannot be the obtaining of pure uranium-235.

But there is another and perhaps more important reason. Without uranium-235 it is impossible to convert the basic mass of uranium (and also thorium) into nuclear fuel. That is precisely why we spoke of potential fuel. As for the isotope uranium-235 itself, that fuel should hold off the onset of an energy famine for a hundred years or so. Therefore, if we want to utilize nuclear fuel for many long centuries, the approach must be quite different.

Nuclear fuel can be produced in the reactor itself, it turns out. We can first of all make plutonium-239, which is obtained from uranium-238, and, secondly, uranium-233, which is obtained from thorium-232. But there is no way of starting without uranium-235.

Reactors that generate energy and at the same time create new fuel are called breeders. It is possible to achieve a situation where the reactor will produce more new fuel than it consumes, that is, the reproduction factor is greater than unity.

To summarize then: technologically feasible ways of utilizing all the supplies of uranium and thorium are available. Consequently, by the most conservative estimates, there is enough fuel to last thousands of years.

And yet adding uranium and thorium to the fuel list does not fundamentally solve the energy problem for mankind; after all, there is a limit to the reserves of mineral wealth in the earth.

Now thermonuclear reactions are quite a different proposition. If we are able to tame the fusion of light nuclei, achieve a controlled self-sustaining thermonuclear reaction, then we can truly say that the energy problem has been resolved. How realistic is such a solution? Just recently, physicists have obtained hydrogen plasma with temperatures of about 60 million degrees. That is the temperature at which a thermonuclear reaction can take place. But the problem is to make such a reaction self-sustaining and to design a fusion (or thermonuclear) reactor. That is what we have yet to do.

There is so much thermonuclear energy in the waters of the oceans that it could satisfy all the power requirements of humanity for a time span exceeding the age of the solar system—a limitless source of energy, indeed.

We thus conclude our talk on fuel. Now let us examine the machines that make the fuel work.

Electric Power Plants

Of course, there are many instances of the direct utilization of energy not in any way connected with the generation of electricity. Gas burning in the kitchen, the launching of a space vehicle (here, the recoil of the products of combustion sends the rocket into space), and even the old steam engines that are occasionally still found. In a number of cases, wind is converted directly into motion and energy.

But in most cases we need electric current—to provide lighting, to power electric engines, to provide for electric traction, to ensure the operation of electric-welding and heating furnaces, to charge storage batteries, and to do many other things. At any rate, today we cannot conceive of delivering energy over distances in any way other than by electric current. It is therefore no exaggeration to say that the principal hero of technology remains the electric power plant.

There are still two basic industrial methods of turning the rotating parts of electric engines, the machines that generate electricity. If this work is done by the energy of falling water, we speak of *hydroelectric power plants*; if the moving force is the pressure of steam on the blades of turbines, we speak of *thermal power plants*.

Included in the class of thermal plants are *atomic power plants*, although they actually differ from ordinary thermal plants in only one way: they use a different type of fuel. But in both cases we obtain heat that is used to produce steam.

The urbanite today is acquainted with the term central heating and power plant. These plants combine a supply of electricity with steam and hot water.

The energy of falling water has been utilized for centuries. The water wheel of the ancient mill is the prototype of the modern hydraulic turbine. A stream of water hits the blades of the wheel and transfers part of its kinetic energy. The blade moves and rotates the wheel.

It is not so easy to position the blades of a wheel in order to obtain the best efficiency. This engineering problem is resolved by specialists in different ways, depending on how the water falls. Naturally, the greater the height (sometimes even 300 metres has been attained) from which the water falls the better the turbine performs.

The latest word in hydraulic-turbine construction are turbines exceeding 500 000-kilowatt capacity. Since power outputs of this kind are generated at rather low rotation speeds (of the order of 100 a minute), the new hydraulic turbines are colossal both in size and weight.

Depending on the direction of flow in the rotor, hydraulic turbines are divided into axial-flow turbines and radial-axial turbines. The Soviet Union has hydroturbines of the radial-axial class with power outputs of 508 megawatts and 7.5-metre rotors.

Hydroelectric power plants generate the cheapest electricity, but they are several times more expensive to construct than thermal power plants and the construction period is much longer. In hydraulic power plants, hydroturbines set in motion hydrogenerators. These are very large synchronous machines mostly with vertical shafts. In such a machine, the diameter of the rotor is 7 to 10 times its length and in some of the largest machines it exceeds 15 metres. This is necessary for the machine to operate stably for variable speeds of the hydroturbine that rotates it. The rotor of a hydrogenerator has a large number of salient poles. For example, the generators of the Dnieper hydropower station have 72 poles. A special direct-current generator (called an exciter) is used to feed the windings of the poles. Hydrogenerators operate at rather low speeds—from 80 to 250 revolutions per minute.

The hydrogenerator of the Krasnoyarsk power station (500 megawatts) operates at 93.8 revolutions per minute, has a 16-metre-diameter rotor and weighs 1640 tons. A 650-megawatt generator is being designed for the Sayano-Shushensky hydropower station.

As I have already pointed out, hydropower generation affects the environment in certain adverse ways. But nevertheless there can be no doubt that hydropower plants have definite advantages over thermal power plants. First of all, hydropower plants do not use up fuel, and we know that fuel is in short supply. Thermal power plants have yet another very great drawback. A considerable portion of the power generated in converting the energy of the fuel into electricity is unavoidably lost.

Still and all, something like four fifths of all the electric power is generated at thermal power plants by means of turbogenerators in which the driving force is the pressure of steam.

To increase the efficiency of the generator, it is necessary to raise the temperature of the steam. This can obviously be attained only by simultaneously increasing the pressure. In modern thermal electric power plants with capacities ranging from 200 to 300 megawatts, the turbines operate at steam parameters of 565 °C and 24 meganewtons per square metre.

But why are high temperatures needed? To put it simply, a steam turbine utilizes the same phenomenon that makes the cover of the tea kettle jump up when the water begins to boil. In other words, in a steam turbine we see thermal energy transformed into mechanical energy, and then from mechanical energy to electricity. Now, in the first transformation (this can be proved rigorously), energy is lost to the extent of a portion equal to the ratio of the ambient temperature to the temperature of the steam.

It is definitely too bad that we have to go through the "thermal stage" in modern devices in order to extract energy. This transition always involves a very great loss of power, and the ideal power plant of the future is one in which energy of any origin can be converted directly into electricity. Since that most important problem is not yet solved, there is only one thing left; and that is to strive for the highest possible temperatures of steam, gas, or plasma.

Difficult as that is, we are still able to achieve efficiencies of about 40% in thermal electric power plants. The steam power generator is an electric machine with a horizontal shaft. The rotor is manufactured together with the ends of the shaft as a single piece of forging made out of special turbo-rotor steel. This is done because the mechanical stresses due to the high speeds of 3000 revolutions per minute reach the very limits of modern materials. For that same reason, the rotor does not have any salient poles. Parts of its cylindrical surface have slots into which the field (excitation) winding is laid. A three-phase alternating-current winding is laid in the slots of the stator.

Because of enormous mechanical stresses, the diameter of the rotor is restricted, and so to obtain sufficient power the machine has to be stretched out in length.

The first Soviet 500-kilowatt turbogenerators were manufactured at the Leningrad "Elektrosila" plant in 1925. In 1964, the "Elektrosila" works turned out a turbogenerator exactly 1000 times more powerful than the first one—500 000 kilowatts.

The desire to obtain more power from a single machine without increasing the already enormous dimensions has led to very complicated designs. For instance, to reduce losses in the windings of the stator, they are made in the form of hollow copper conductors in which water constantly flows. The excitation winding is cooled with hydrogen under a pressure of about 4 atmospheres. The use of hydrogen, which is 14 times less than the density of air, made it possible to increase the power of the turbogenerators by 15% to 20%.

The five-year plan of 1976-1980 calls for the manufacture of turbogenerators of between 1000 and 1200 thousand kilowatts for thermal and atomic power plants.

One of the most interesting power plants in the world has been built in the Soviet Union. It is called U-25 and feeds about 7000 kilowatts of electricity into the electric network. This is the world's largest plant for generating electricity by the method of magnetohydrodynamics, MHD for short. An MHD-generator has no rotating parts.

The idea behind the operation of this interesting generator is extremely simple. A flux of ions having considerable kinetic energy (this is, a jet of plasma) passes through a magnetic field across the magnetic lines of force. The ions are acted upon by the Lorentz force. Recall that the intensity of an induced electric field is proportional to the rate of ion flux and to the magnitude of magnetic induction. The electromotive force is perpendicular to the motion of the ions. That is the direction in which the current arises and closes on an external load. The electrodes that receive the current are in direct contact with the plasma.

Electric energy is generated due to the fall in energy of the plasma jet. An MHD-generator permits raising the efficiency of the power plant to 60% and more.

The critical factor in generating cheap power from MHD is the magnetic field in the channel. This must be an intense field. An ordinary electromagnet with a copper coil can produce such a field but the magnet is then unwieldy, complex in design, and expensive. What is more, it consumes a good deal of power itself. In this connection, designers have developed a new concept in designing magnets with a superconducting winding. This kind of magnet can generate the necessary magnetic field of required intensity and with low power consumption and insignificant heating. Calculations have shown that the considerable expenses needed to obtain temperatures close to absolute zero justify themselves.

From this brief summary we see that traditional ways of increasing power generation have not yet exhausted themselves. However, one can hardly expect progress only along that road.

Aside from the fact that fuel supplies and opportunities for utilizing hydraulic sources of energy are coming to an end, one should not lose sight of the appreciable effect on the environment of the construction of new electric power plants. Ecologists have warned of the necessity of a very cautious approach to interfering in the life of rivers. Power engineers are reminded of the enormous quantities of ash that are thrown into the atmosphere in the course of fuel combustion. In one year, the earth's atmosphere receives 150 million tons of ash and about 100 million tons of sulphur. Particularly disquieting is the increase in the atmosphere of carbon dioxide. Every year, 20 thousand million tons of it are added to the atmosphere. During the past 100 years, the quantity of carbon dioxide in the atmosphere has increased 14%.

There are two reasons for this growth: the destruction of vegetation on the earth and, what is most important, the release into the atmosphere of this "gaseous ash" in the combustion of ordinary fuel. This ceaseless increase can lead to ruinous consequences, of which the worst is an increase in the temperature of the atmosphere by 1.5 to 3 degrees. This might seem to be a small rise but in turn it can lead to an irreversible melting of the polar ice caps. Climatologists claim that the ultimate permissible increase in carbon dioxide in the earth's atmosphere cannot exceed several tens percent.

Nuclear Reactors

As has already been pointed out, atomic power plants belong to the class of thermal electric power plants. The difference lies in the production of steam that is then directed onto the blades of a turbine. A nuclear reactor could be called a nuclear boiler because of the similarity of power generation.

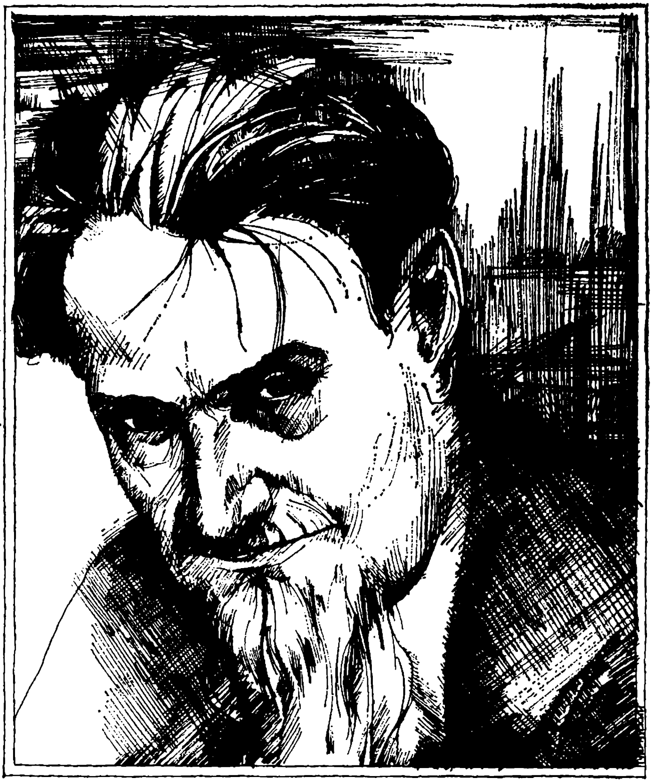
A nuclear reactor is ordinarily in the form of a cylindrical building. The walls have to be very thick and made of materials that absorb neutrons and gamma radiation. Reactors generating something like 1000 megawatts of electricity come in a variety of sizes depending on the fuel used, the method of moderating neutrons, and the mode of heat removal. But in all cases the size is impressive: the height of a 5- to 10-storey building and some ten or so metres in diameter.

Nuclear power engineering began to develop right after the Second World War. In the Soviet Union, the most important nuclear investigations were headed by Igor V. Kurchatov, a marvelous scientist and an excellent organizer.

In the Soviet Union and in other countries, a number of designs were tested. The first stumbling block is to decide the isotopic composition of the uranium or other nuclear fuel used. The engineer is then confronted with the form in which the fuel is supplied as a solution of uranium salts or in solid chunks. Solid fuel elements can come in a variety of shapes. Bars can be used but long rods appear to be most suitable. A very essential role is played by the geometry of the fuel elements (the positions they are placed in). Engineering calculations are needed to find the most suitable positions of the control rods that absorb neutrons. Their movements (all automatic, naturally) must ensure the required value of the neutron multiplication factor.

Differences in the behaviour of slow (thermal) neutrons and fast neutrons permit dividing reactors into two categories, namely, reactors with a neutron moderator and breeder reactors.

A reactor designed to operate on the principle of neutron moderation can use natural uranium. The amount of moderator has to be such as to prohibit large numbers of neutrons from being absorbed by uranium-238 nuclei. Now there are approximately 140 times more of these nuclei than there are uranium-235 nuclei. If there is not



Igor Vasilievich Kurchatov (1902-1960)—prominent Soviet physicist and a remarkable organizer. He headed the work of the atomic project in the Soviet Union. He began his scientific career in the field of solid state physics and set up the theory of seignette-electrics (also known as ferroelectrics). At the beginning of the 1930s he began research in the physics of the atomic nucleus. Under enough moderator, the neutrons will not diminish their speeds to the thermal level and will be absorbed by uranium-238 nuclei, and the chain reaction will come to a halt. A reactor operating on natural uranium or uranium slightly enriched in uranium-235 will still create a new fuel, plutonium. But much less plutonium will be produced than the uranium nuclei that are "burned" up.

Up until just recently, most atomic power plants used thermal-neutron reactors. There are four types of such reactors: water-water reactors with ordinary water as the moderator and heat-transfer agent (coolant); graphite-water reactors with water coolant and graphite moderator; reactors in which heavy water is the moderator and ordinary water is the coolant; and, finally, graphite-gas-cooled reactors.

The reason why specialists in the field of atomic power engineering concentrated their attention on reactors operating with thermal neutrons is apparently due to the fact that it is extremely difficult to enrich uranium in the 235 isotope. The reader will recall the remark made earlier that when only the isotope uranium-235 is used, enormous reserves of potential nuclear fuel are left untouched.

At the present time there is a tendency to have nuclear reactors operate on highly enriched fuel with no neutron moderator.

Suppose the mixture in the reactor consists of one part of uranium-235 and one part of uranium-238. In that case, the number of neutrons that fall out of the chain reaction due to capture by uranium-238 may exceed the number of neutrons fissioning nuclei of uranium-235 and sustaining

---

his guidance, important studies were carried out in the field of nuclear isometry, the resonance absorption of neutrons, and artificial radioactivity.

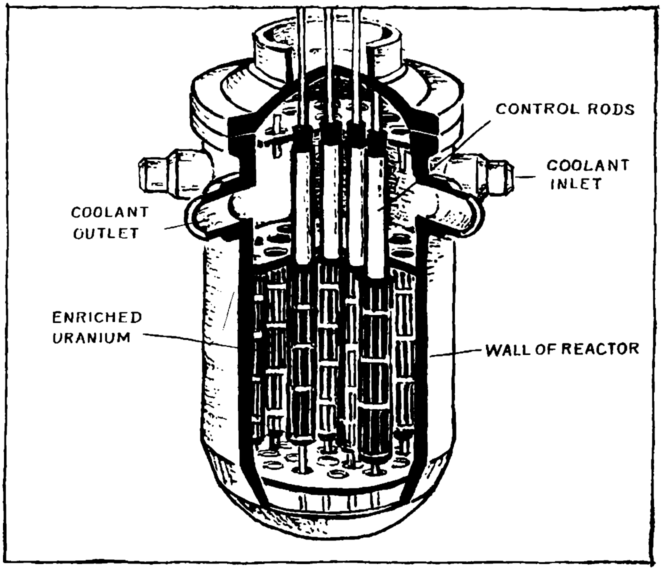


Figure 6.2

the chain reaction. Such a reactor is termed a *breeder*. Depending on the geometry of the rods or bricks of nuclear active and potential fuel, it is possible to design breeder reactors with a great variety of percentage ratios of the two types of fuel and with different reproduction factors.

Two examples will suffice to give the reader some idea of the parameters of nuclear reactors.

Figure 6.2 is a general view of the design of a nuclear reactor used in submarines of the United States. The coolant here is ordinary water. Since ordinary water captures neutrons roughly 600 times more effectively than heavy water, such a reactor can operate only on uranium-

238 enriched in uranium-235. Instead of the natural portion of 0.72% in the fuel of these reactors we find from one to four percent of uranium-235. The reactor is capable of generating 1100 megawatts of electricity, is about 5 metres in diameter, has a height of 15 metres (a 5-storey building) and wall thickness of about 30 centimetres. If such a reactor is charged with 80 tons of uranium oxide containing 3.2% of uranium-235, it will operate 10 to 12 months (and then the rods have to be replaced). The water in the reactor heats up to 320 °C and circulates under a pressure of about 300 atmospheres. The hot water turns to steam and is delivered to the blades of a turbine.

We now give a brief description of the design of a powerful breeder reactor to be built in France. It is called Superphénix.

The fuel will be a mixture of plutonium-239 and uranium-238. No moderator will be needed and so the neutrons do not lose speed from the time of generation during nuclear fission up to an encounter with another atomic nuclei of the fuel.

The fact that the reactor employs fast neutrons makes for a high degree of compactness. The core of the reactor will not exceed 10 cubic metres. Thus, there will be a large liberation of heat per unit volume.

Heat removal cannot be done with water because water moderates neutrons. Liquid sodium will be used. Sodium melts at 98 °C and boils at 882 °C at atmospheric pressure. For technical reasons, the temperature of liquid sodium must not exceed 550 °C. There is therefore no need to raise the pressure of the coolant (this is resorted to when the coolant is water).

The Superphénix reactor will have the following dimensions: inner diameter 64 metres, height about 80 metres. This is a big 20-storey building! The core of the reactor is in the form of a hexagonal prism made up (like a bundle of pencils) of thin rods each 5.4 metres in length. The fuel rods alternate with control rods.

There is no need (nor space) here to describe how the core of the reactor is cooled. Suffice it to say that it is done in three stages. A first loop contains sodium; it removes the heat from the reactor and delivers it to the boiler, where it is taken up by a second sodium loop. In the second heat exchanger, the heat is taken up by a third loop with a circulating water-steam mixture. Then follows the ordinary route to the steam turbine.

Calculations show that this setup should generate 3000 megawatts of thermal capacity and 1240 of electricity.

I cannot help stressing once again that the necessity of converting nuclear energy into electric energy passes through the thermal stage, leaving a poignant feeling of disappointment. It is something like installing an automobile engine on a cart. But so far no one has come up with a better idea of how to bypass that stage, which creates perhaps the greatest difficulties in the construction of atomic power plants. Besides the common drawback of all thermal power plants, we have here the added necessity of including intermediate pipelines. This is done because we have to eliminate the unacceptable radioactivity of the water vapour entering the turbine.

Here are some more figures of this design. The maximum neutron flux per square centimetre per second is equal to 6.2×10^{15}. The reproduction factor will be equal to 1.24. Used up fuel elements will be replaced once a year. In the first loop, there will be a flow of liquid sodium (technically called the mass flow rate) of 16.4 tons a second. Superheated steam will emerge under a pressure of 180 bars and at a temperature of 490 °C.

A few words are in order about the "ashes" of nuclear fuel. A great number of radioactive isotopes appear in the fission of fuel nuclei. This is an uncontrollable process, but we have the opportunity of obtaining any isotopes we want merely by placing substances in the reactor. Such substances absorb neutrons and create new atoms.

Of course, radioisotopes can be obtained in accelerators by subjecting various materials to bombardment with protons or the nuclei of other elements.

The number of artificial elements obtained to date is very great. The "empty" places in the periodic table of elements have been filled up: elements No. 61, 85, and 87 do not have long-lived stable isotopes, and so there are none in nature. The Mendeleev table of elements has been extended. Elements above 92 are termed transuranium elements. Each of the transuranium elements, all the way up to element 105, has been obtained in several isotopic variants.

Also, in addition to new chemical elements, large numbers of radioisotopes have been obtained of those chemical elements that are encountered in the earth's crust in their stable form.

Radioisotopes have been widely used for many years in the sterilization of food products by gamma rays, in flaw detection, and as generators of electric power that use electrons produced in decay processes. This list could be extended considerably.

The usefulness of radioactive isotopes is, unfortunately, commensurate with the trouble engineers have in protecting persons that deal with radioactive radiation.

Nuclear fuel waste contains 450 kinds of atoms, including uranium-237 and neptunium-239, which convert into neptunium-237 and plutonium-239.

Unlike coal or petroleum, nuclear fuel does not burn up completely. In a number of cases, nuclear reactors operate with enriched fuel containing between 2.5% and 3.5% of uranium-235. At some point, a reactor ceases to deliver energy because in the fission process a large number of isotopes are produced that capture neutrons and prevent the fission process from developing. When a reactor is shut down, there is roughly 1% of uranium-235 and a somewhat smaller amount of plutonium-239 left.

There can be no question of throwing away such ash containing large quantities of useful fuel. A big chemical works could be linked up with an atomic power plant. It would have to be completely automated since it would be handling highly radioactive materials. Special measures would also have to be taken to protect personnel from gamma radiation.

At these plants, used fuel elements would be ground up and dissolved. The pure fuel (uranium and plutonium) could be extracted and returned for the manufacture of fresh fuel elements.

There still remains appreciable quantities of radioactive waste in the form of "hot" (radioactive) solutions that have to be buried. And what has to be ensured is that for many centuries the disposal sites will not be touched in any way.

Specialists are more or less optimistic. It is believed that the storing of cans (drums or barrels) of radioactive solutions up to one kilometre underground in specially selected spots should guarantee one hundred percent safety. What kind of sites are suitable? That is a matter for geologists to decide. Only places safe from earthquakes, of course. Also, one must be sure there are no subterranean rivers or streams. Salt deposits have been found to satisfy such conditions. And, of course, the barrels cannot simply be dropped into a kilometre-deep well. To ensure dissipation of heat that the barrels would be emitting constantly, they will have to be dispersed at distances of at least up to 10 metres apart.

Thermonuclear Energy

We have already mentioned that chemical and nuclear reactions have much in common. Since heat is generated not only in reactions of decomposition but also, frequently, in the combining of two molecules into one, we can expect that atomic nuclei behave in a similar manner.

It is not difficult to see what fusion reactions (combining of nuclei) would be advantageous from the energy point of view if we know the masses of the atomic nuclei.

The deuterium nucleus has a mass of 2.0146 mass units. If two nuclei fuse into one, we get ^{4}He, which has a mass of 4.0038 instead of $2 \times 2.0146 = 4.0292$. The excess mass of 0.0254 is equivalent to roughly 25 MeV of energy, or 4×10^{-11} joule. One gram of deuterium contains 0.3×10^{24} atoms. Now if such a reaction took place, then two grams would yield 10^{13} joules of energy! It turns that the most promising are the fusion reactions of the heavy isotopes of hydrogen: deuterium, tritium. Even ordinary hydrogen can be put to use as thermonuclear fuel.

It is thus clear that atomic energy can be generated in two forms: the splitting up of heavy atomic nuclei in so-called fission reactions and the combining of light atomic nuclei in so-called fusion reactions, which are also called thermonuclear reactions.

Fusion reactions, if harnessed, could ensure our power needs for millions of years to come (and the level of the world's oceans would not go down perceptibly in the process of using the water). It would indeed be a limitless ocean of power, and at very low costs—practically free of charge.

However, it is no easy job to harness this "free" source of power. The point is that all atomic nuclei are positively charged. That means an enormous amount of energy is required to bring the nuclei close together.

The only way to obtain this energy is to convert a substance into the plasma state, which means stripping atomic nuclei of their clouds of electrons and then raising the temperature of the plasma to a point where the nuclei begin to collide (collisions occur at distances of 10^{-13} centimetre apart) despite the electric repulsive forces.

Calculations turn out to be quite disappointing. The reader can himself try to calculate the amount of energy of electrostatic repulsion using the formula e^2/r and then check the needed temperatures (you will have to recall the formula that relates the temperature and the kinetic energy of any particle). The answer is in the tens of millions of degrees.

To summarize, then, we need to produce a high-temperature plasma. There are two ways to do this: one has been in use for over two decades, the other is fifteen years or so younger.

The former method consists in creating a thermonuclear reactor by "driving" the plasma into a magnetic bottle (that's what it's called).

If a magnetic field is imposed on a gas-discharge tube and is made to coincide in direction with the electric field, a plasma column is generated. We know that the charged particles of the plasma will describe helical trajectories. We can say that the particles move in a single ring surface current. The stronger the magnetic field the smaller the radius of the plasma column. The force acting on the current of charged particles via the magnetic field is what produces the column, which does not come in contact with the walls of the gas-discharge tube.

In principle, we have thus produced plasma that actually "hangs in the air".

Calculations show that if the initial hydrogen pressure is of the order of 0.1 millimetre of mercury, the radius of the column is 10 centimetres, and the intensity of the discharge current is 500 000 amperes, then the temperature of the plasma should be sufficient for a fusion reaction to start up.

However, there are very many obstacles still in the way of a controlled thermonuclear (fusion) reaction. The plasma column is the main stumbling block. For a number of reasons it is highly unstable and disintegrates almost instantaneously. The problem gets under control only if one is able to make a magnetic bottle with what might be called "feedback": that is, such that the random fluctuations that disrupt the column give rise to counteracting forces capable of restoring the column.

In the middle of 1978, a group of American physicists working at Princeton University succeeded in heating plasma to 60 million degrees. This success was achieved by using magnetic bottles called Tókamak (we discussed them in book three of this series). Tokamak—a Soviet device—is an acronym that stands for toroidal, chamber, magnet. The temperature attained is sufficient for the fusion of nuclei of deuterium and tritium. Since then, Soviet and also American scientists have attained still higher temperatures and better parameters for other aspects of this extremely complicated undertaking.

There is still a big problem, however, and that is to confine the hot plasma in a restricted volume for sufficiently long time intervals. Ways of doing this are not so obvious to the engineer. It very well may be that the achievement of controlled fusion reactions is an extremely costly business. But research goes on and scientists remain optimistic.

Attempts have recently been made to attain controlled thermonuclear fusion by means of laser emission. Lasers are now in operation with power outputs of about 10^{12} watts. This power can be delivered to a substance we wish to convert to plasma in the form of pulses of light of duration 10^{-9} to 10^{-10} second. Naturally, when light of such colossal power strikes a solid, the substance is in-

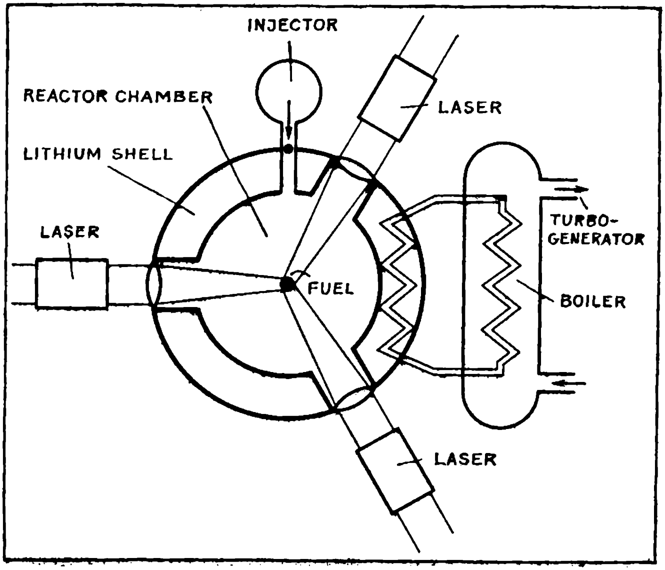


Figure 6.3

stantaneously ionized and passes into the plasma state. What scientists are aiming at is a situation in which a deuterium-tritium plasma is created with a temperature of 10^8 degrees, the temperature being maintained for a time interval sufficient for a chain reaction to develop. This requires plasma of sufficiently high density to ensure a large enough number of collisions of the nuclei.

Those are the underlying ideas of the reactor sketched in Figure 6.3. A solid (frozen) pellet consisting of hydrogen isotopes falls inside a vessel that has been evacuated to a deep vacuum. When the pellet passes through the centre of the vessel, powerful lasers are switched on that transform the solid into a state of plasma. For the reactor to start up, the energy liberated in the time interval between the start and finish of the reaction must be sufficient to maintain the temperature required to continue the reaction. Calculations show that the plasma must have a density from 10^3 to 10^4 times that of the solid, or about 10^{26} particles to every cubic centimetre. A laser is capable of producing the needed compression.

In principle, it is possible to obtain the requisite temperature and density. Then what happens? The fusion energy of the nuclei is conveyed to the neutrons that are released in the reaction. These neutrons fall on the lithium shell of the vessel. The lithium, via a heat exchanger, transfers the energy to a turbogenerator. Part of the neutrons react with the lithium and produce tritium, which is needed as fuel.

The principle is simple but the end is still a long way off, and it is quite possible that fresh and unexpected obstacles will crop up. It is very hard to foresee the requirements of such a reactor when power generation actually begins. Research workers are convinced that producing such high power outputs in small volumes of substance will reveal entirely new phenomena.

Solar Energy

The conversion of solar energy to electricity by means of photocells has been known for a long time, but only recently has the idea occurred that this phenomenon could be used as the basis for an electric power plant. The suggestion might appear at first glance to be rather wild. Figure it this way: to make a power plant of 1000 megawatts, you would have to cover an area of 6×6 square kilometres with *solar cells* (these are special photocells for the conversion of solar energy into electricity). And the calculations are for the sun-drenched Sahara desert. What would you say of a plant like that for the middle

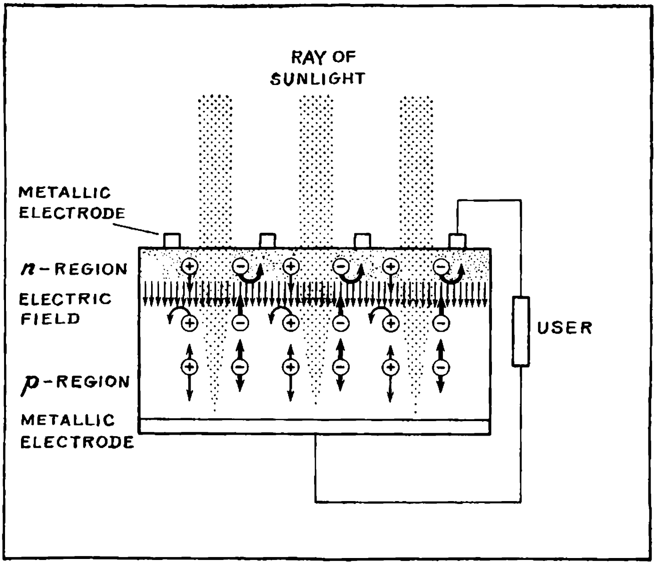


Figure 6.4

of Europe? Here there are not so many sunny days and the area would have to be doubled at the very least. Science fiction, pure and simple, says the reader. Just imagine what a power plant like that would cost to build!

True enough. But compare that with the advantages of such a mode of power generation. We do not use up any material of the earth, we do not pollute the environment with any kind of waste. Aren't those two factors strong enough inducements to get down to a serious study of ways of building cheaper solar cells, optimal arrangements of the cells, and the focussing of solar radiation? Many scientists (and I am one of them) are convinced not only that the problem deserves our closest study but hope that precisely this path will lead to electric power plants of the future. In a few years that might very well be problem No. 1.

Is not this optimism rather premature? What is the situation today in this field? First of all, what kind of solar cells does industry offer at the present time?

Figure 6.4 is a diagrammatic sketch of how solar energy is converted into electricity. The cell consists of a semiconductor *p-n* layer squeezed between metallic electrodes. The sun's rays create free electrons and holes, which are sent in opposite directions by means of a contact voltage, thus generating a current.

There are three main types of such cells. Homocontact cells in which a *p-n* sandwich is created by alloying silicon. A diffusion process is used to make a thin (0.3 micrometre) *n*-layer and a relatively thick (300 micrometres) *p*-layer. Heterocontact cells consist of two different semiconductors. An *n*-layer of cadmium sulphide is sprayed onto a metallic backing to a thickness of 20-30 micrometres; then on this surface, chemical methods are used to create a *p*-layer of cuprous sulphide 0.5 micrometre in thickness. A third type of cell makes use of the contact voltage between gallium arsenide and a metal separated by an extremely thin (20 angstroms!) dielectric film.

For optimal utilization of the energy of the whole solar spectrum, semiconductors with electron binding energies of about 1.5 electron volts are suitable. In principle, efficiencies of 28% are attainable in solar cells.

Silicon homocontact cells, which have a number of technical advantages and have been studied in the most detail, have efficiencies from 11% to 15%. Silicon solar cells have been in use for more than twenty years. The material here is quartz sand (silicon dioxide) from which pure silicon is obtained. Single crystals are manufactured with thicknesses of about 0.3 millimetre in the shape of circular disks. In recent years, a process has been developed for manufacturing monocrystalline tape. The technology of semiconductor doping (the addition of impurities to a material) is well developed and permits establishing a *p*-layer in the silicon disk. In order to reduce the reflection of sunlight from the silicon, the surface is covered with a thin film of titanium oxide. With a light intensity of 100 milliwatts per square centimetre, a disk sets up a voltage of 0.6 volt. The current density of a short circuit is equal to 34 milliamperes per square centimetre. There are variety of ways of collecting the cells into batteries. Silicon monocrystalline disks are now being manufactured with diameters from 5 to 7.5 centimetres. They are fixed between plates of glass and can be combined to form a rather powerful source of electric current.

It is hoped that new technological processes will develop cells with a much larger area.

The chief obstacle today in the way of using solar cells for industrial power generation is the high cost of manufacturing high-quality monocrystalline tape.

There is considerable hope of making solar cells out of thin polycrystalline layers. The process itself will be cheap but the efficiency will fall drastically. The search is on for cheap methods of obtaining effective solar cells.

At the same time, researchers are seeking ways of increasing the energy falling on a single cell.

There are power plant designs consisting of 34 thousand mirrors reflecting sunlight and directing the radiation to receivers located at the top of a tower 300 metres high.

If we employ concentrated solar energy, we have to see that the increase in the temperature of the cell does not greatly affect the overall efficiency. In this respect, gallium arsenide cells have certain advantages.

Projects are now being considered of siting solar power plants on the top of high mountains where good condi-

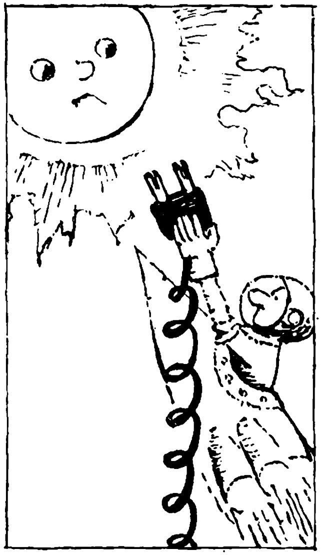


Figure 6.5

tions of solar illumination are ensured. We also have a detailed design of a power plant to be flown on an artificial earth satellite. Such space-borne power plants can receive the energy of solar radiation without any losses and then in the form of microwaves deliver it to the earth where it can be transformed into electricity. And this is not science fiction. Engineers are seriously considering a satellite-borne power plant 25 by 5 kilometres in size. This area could accommodate 14 000 million photocells! The plant would weigh 100 000 tons and would generate the equivalent output of several tens of the largest atomic power plants now operating; that is, of the order of ten thousand megawatts.

Detailed projects have already been worked out and small models are undergoing tests.

Power from the Wind

The air masses of the earth's atmosphere are in constant motion. Cyclones, storms, the constant trade winds of the tropics, and light breezes are only some of the great diversity of energy manifestations of streams of air. The energy of the wind has for ages been used in sailing boats and to drive windmills. The overall mean annual power developed by all air streams over the entire globe comes to the fantastic figure of 100 000 million kilowatts.

Meteorologists are well informed about wind speeds in different places on the globe and at different altitudes above the earth's surface. The wind is a temperamental thing and so all estimates involve a mean velocity of 4 metres per second and a mean altitude of 90 metres, which is a modest figure for a coastal strip.

Apparently, the most suitable sites for utilizing the "blue" energy of the wind are coastal areas of seas and oceans. It turns out that if Great Britain took advantage of the wind even in rather calm weather (Britain is the richest in wind among the European countries), it could generate roughly six times that produced by all the electric power plants of the country at present. In Ireland, the energy of the winds exceeds electric consumption of the whole country by a factor of one hundred (true, perhaps it is not that there is so much wind but rather so few power plants).

Some twenty or so years ago wind did not rate very high as a source of future power. But the trends of power engineering have been changing before our very eyes. Commissions are set up one after the other to look into the possibilities of new and less expensive sources of energy. The resources of the earth are now regarded in a different light: humanity has begun to think about what is reasonable and what is not as far as utilization of the natural resources hidden in the earth goes. That is why the power

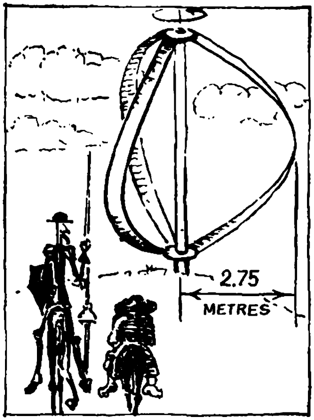


Figure 6.6

of the wind is now being taken seriously. Taking a sober view of engineering possibilities, we may regard as realistic the harnessing of a fraction of one percent of the 100 000 million kilowatts of available wind power. Even that is a good deal.

Enormous "windmills" have already been designed with vanes reaching to 100 metres and more and towers of about that height; windmill vane tips have been calculated to reach speeds of about 500 kilometres per hour. In ordinary weather, a windmill of that class could be expected to have a power capacity of 1 to 3 megawatts. Several thousand such windmills operating in a country with frequent strong winds could supply all the electricity needed. In Western Europe, 1261.6×10^{12} watt-hours of electric power were generated in 1973. In principle (if one does not skimp on initial outlay) that is but a small part of the energy that can be feasibly extracted from the wind! Gigantic windmills are already under construction.

Calculations show that a wind-driven motor has a maximum output when the rotor reduces the wind speed by one third. One should not think that wind-driven units must always imitate windmills. Rotors with a vertical axis of rotation are quite possible. The rotor shown in Figure 6.6 is capable of power outputs of the order of 20 kilowatts. The advantage of such a rotor is its independence of the direction of the wind; a drawback is that it operates only when there is a strong wind. Rotors of this kind are manufactured with diameters of up to 5.5 metres.

Quite naturally, generators driven by wind must be positioned on a relatively small area, but still at distances apart so that their interaction does not play any role. A power plant with a capacity of 1000 megawatts requires an area of about 5 to 10 square kilometres.

7. The Physics of the Universe

Measuring Distances to the Stars

Today it is no longer possible to draw a strict boundary line between astronomy and physics. As long as astronomers, like geographers, confined themselves to descriptions of the stellar sky, the subject of astronomy drew small attention from physicists. However, the situation changed radically a few decades ago, particularly after observations began to be made from artificial earth satellites and the moon.

Without the terrestrial atmosphere as a hindrance, it is possible to receive all signals coming to us from the far corners of the universe. These include fluxes of a variety of particles and electromagnetic radiation of practically the whole spectrum, from gamma rays to radio waves. And opportunities for observing the stars in the visible portion of the spectrum have increased immeasurably.

Now the study of particle fluxes and radiation of the electromagnetic spectrum most definitely belong in the sphere of physics. If we also note that studying outer space brings us face to face with a great diversity of phenomena that do not as yet bow to a unique interpretation and if we further note that we can and must be prepared for the physics of the universe leading to the discovery of entirely new laws of nature, it becomes clear why explorers of the stellar world are today mostly

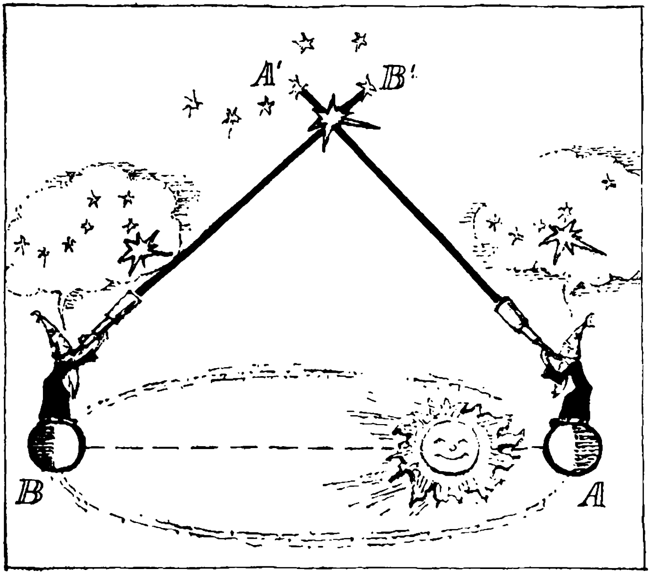


Figure 7.1

physicists—physicists by education and mode of thought.

We begin our discussion of the universe with a classical problem of astronomy. Measuring distances from the earth to distant stars. At the present time, the distances between the earth and the sun or the earth and the planets are measured to a high degree of accuracy by means of radar. The mean distance to the sun is equal to 149 573 000 kilometres. It has been calculated to within a millionth fraction of the distance.

But astronomers had already measured distances inside the solar system without the aid of radar, using a simple (in principle) method called *triangulation*.

Consider a group of stars, from two different spots, which should be as far apart as possible. This separation is termed the *base* by astronomers. As sketched in Figure 7.1, astronomers first used very simple devices and, later, excellent telescopes to measure the angles between the directions to individual stars. They noticed that it is possible to choose a group of stars that moves across the sky as an integral whole. No matter what positions the stars are observed from, the angles between the directions to them remain the same. But among such stars they often found some one star that clearly moved with respect to its neighbours. Taking one of the "fixed" stars as a point of departure, it was possible to measure the angular displacement of the star that moved relative to the fixed constellation. That angle received the name *parallax.*

As far back as the seventeenth century, after Galileo's invention of the telescope, astronomers had already measured the parallaxes of the planets by observing their displacements with respect to the "fixed" stars. It was then calculated that the earth is at a distance of 140 million kilometres from the sun. Very accurate indeed!

To the naked eye, the mutual positions of the stars remain unchanged. Only by taking photographs of the stars from different positions is it possible to detect parallactic displacements. The largest base for such a measurement is the diameter of the earth's orbit. If we take two photographs of some portion of the sky from one and the same observatory with a time interval of half a year, the base will be equal to nearly 300 million kilometres.

Radar cannot be used to measure stellar distances and so the diagram shown in Figure 7.1 is quite modern.

This suggests that there are stars which are in perceptible motion relative to other stars. But it would be extremely illogical to assume that there are fixed stars and moving stars. The conclusion we are forced to is that the stars whose mutual positions remain unchanged are much farther away than a "wandering" star (or planet). Be that as it may, with good instruments we are in a position to measure the parallaxes of many stars.

Parallax measurements to within a hundredth of a second of arc have been carried out for many stars. It turned out that the stars closest to us lie at distances greater than one parsec.

A *parsec* is the distance that yields an angular displacement of one second if the base is the diameter of the earth's orbit. It is easy to calculate that one parsec is equal to 30 million million kilometres.

Astrophysicists also use what is known as the light year. The *light year* is the distance traversed by light in one year. One parsec is equal to about 3.3 light years. Light travelling at 300 000 kilometres a second takes half a day to cross the whole solar system. We now come to an important and amazing conclusion: our planetary system is very much alone in the universe.

The parallax method can be applied to distances of the order of hundreds of light years. We then have the problem of how to measure distances to more distant stars. This turns out to be not at all simple, and any confidence in the correctness of approximate estimates (for the most part we can only be sure of one significant digit) is obtained by correlating results of different measurements.

At any rate, the decisive factor in determining distances to faraway stars is that there are many so-called double stars among the millions upon millions of stars in our galaxy. Double stars also go by the name variable. The reason for this is clear: the brightness of the star varies because of the mutual motions of the two stars of the pair. The period of variation can be measured precisely.

Suppose we are observing a very distant cluster of stars. We are justified in concluding that the differences in observed brightness are not related to any difference in distance. We of course know about the inverse-square law (that is, the intensity of any radiation decreases with the inverse square of the distance), we also know about brightness diminishing if light rays pass through accumulations of stellar gas. But if a small patch of the night sky has been chosen and the stars are far away from us, they can be regarded as being in identical conditions.

Studies of double stars lying in the Small Magellanic Cloud led to the conclusion that there is a relationship between the period of a double star and its luminosity. This means that if certain stars are at approximately the same distance from us and the period of variation of brightness (this period is usually between 2 and 40 days) is the same, the stars will appear to be equally bright.

Having established this rule and being convinced of its usefulness, we can utilize it for stars that belong to different clusters. By comparing stars whose periods of variation are the same but whose brightnesses are different, we can assert that they lie at different distances from us. In order to determine how great these distances are, we can use the inverse-square law. But this method permits determining only the relative distances of variable stars from the earth. We can apply it only to very distant stars and cannot use it to calibrate our information about the absolute stellar-earth distances obtained in measuring parallactic displacements.

The scale for measuring distances between very distant stars was found by employing the Doppler effect.

The formulas that we discussed on page 195 of the third book of this series hold true for any vibrations. Therefore the frequencies of spectral lines observed in the spectrum of a star permit determining the velocity of the star away from the earth or towards the earth. Since in the equation

$$\nu' = \nu \left(1 \pm \frac{v}{c}\right)$$

c is the velocity of light (300 000 km/s), it is clear that the star must have a high velocity and the spectrograph must be of extremely high quality for us to detect a displacement of the spectral lines.

Please note that the scientist is quite certain that the hydrogen on the star (that is, the hydrogen that has signalled to us) at an unimaginably colossal distance from us is the very same hydrogen that we deal with here on the earth. If the star were at rest, the hydrogen spectrum would have to be exactly like the spectrum we obtain from a gas-discharge tube (that is how confident the physicist is of the unity of the world!). But the lines are noticeably shifted and stellar velocities are in the hundreds and even tens of thousands of kilometres a second. No one has any doubts about this explanation. How could there be? The spectrum of hydrogen consists of a very large number of lines, and we see the displacement of all lines (not just one) of the spectrum in perfect accord with the Doppler equation.

But let us return to the measuring of stellar distances. Of what help can our knowledge of stellar velocities be? It is very simple. Provided of course that you have observed the star to have moved a certain distance during the year (all this of course with respect to other stars, which in this case may be regarded as "fixed"). If we know the arc displacement φ of a star (φ being perpendicular to the light ray that reaches us), then, knowing the tangential velocity, we can find the distance to the star R from the formula

$$\frac{R\varphi}{t} = v$$

In place of t we can substitute the time of translation of the star.

But, says the reader, the formula involves the tangential velocity, whereas we do not know the direction of motion of the star. Quite true, and that is why we have to take the following approach. A large number of double stars with the same blinking period are selected. Then the radial velocities of all these stars are measured. The radial velocity will vary from zero (if the star is moving at right angles to the ray of light) to a maximum (when the star is moving along the ray). If we assume that, on the average, the tangential and radial velocities are the same, we can put the mean values of the measured velocities into the above formula.

The Expanding Universe

Having measured stellar distances, we can now describe the world of stars as follows. The observable portion of the universe can be broken up into an enormous number of aggregations of stars called *galaxies*. Our solar system lies in the Galaxy (our galaxy). It is called the Milky Way and can be seen in the night sky as a large patch of grey. The Milky Way Galaxy is in the form of a disk of thickness about 100 000 light years. It contains about 10^{11} stars of different types. The sun is one such star and lies on the outskirts of the Galaxy. The component stars are so far apart that they never collide. On the average, interstellar distances are 10 million times greater than the sizes of the stars themselves. To obtain a similar rarefaction in the terrestrial air space we would have to reduce the density of the air by a factor of 10^{18}.

Turning now to the mutual positions of different galaxies, we find the situation is quite different. The mean distances between galaxies are only several times greater than the sizes of the galaxies themselves.

Astrophysicists have learned a great deal about the peculiarities of mutual motions of stars belonging to a single galaxy. All this is beyond the scope of our story.

But in a book devoted even to the very fundamentals of physics we cannot help mentioning an exceptionally important observation. It has been established with great certainty, in studies of the Doppler effect in spectra belonging to stars of different galaxies, that all galaxies are racing away from us in all directions. What is more, it has been shown that the velocity of recession of a galaxy is directly proportional to its distance from the earth. The most distant visible galaxies are racing away from us with velocities of recession close to half the velocity of light.

When I say "from us", it sounds strange—as if God made the earth and strewed stars about in the surrounding space. That was the picture of the world in ancient times (Aristotle) and in the Middle Ages. The universe had boundaries beyond which lay the empyrean of ancient belief—the abode of God.

To modern man, the universe is not visualized as having any boundaries. For if there is a boundary, what lies beyond? We must picture the universe as being without bounds. On the other hand, one certainly cannot believe that the earth and sun are special bodies in the universe. This would contradict every piece of knowledge acquired by astrophysicists. But the galaxies are racing away from us! How can we reconcile this fact with our requirements for a model of the universe in which there are no boundaries, matter is more or less uniformly distributed, and the picture of the universe as seen by an inhabitant of any star is the same?

The intellectual necessity of the existence of such a model led Einstein to the following fundamental conclusion. Euclidean geometry that is so useful in our everyday life does not hold true when we deal with the unimaginably great distances encountered in our studies of the stellar world. A rejection of Euclidean geometry leads to a rejection of pictorial models of the universe.

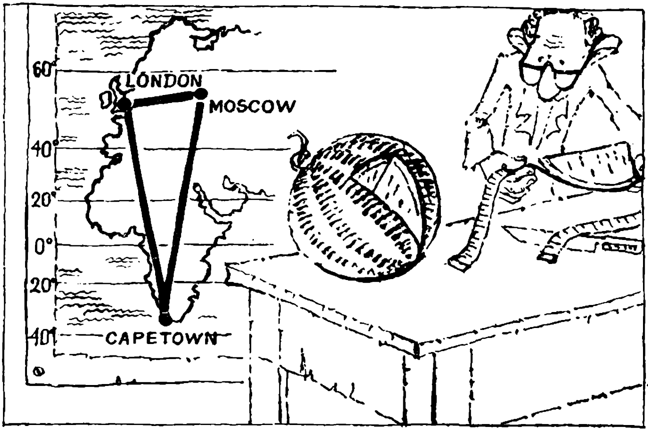


Figure 7.2

No matter, this is not the first time that we have had to give up pictorial conceptions of the world.

Taking leave of Euclidean geometry, we can propose a model of the universe that is simultaneously closed and yet does not have any centre or any boundaries. In such a model, all points of space are of equal status.

At first glance, it would seem that Einstein is asking for a great deal. We are so used to two parallel lines never meeting, to the sum of the squares on the sides of a right triangle being equal to the square on the hypotenuse. Yes, but let me remind you of certain lessons in geography.

On the globe of the earth, the parallels of latitude are parallel lines. But on a map? What kind of map? you may ask. Geographical maps are constructed in different ways. If we depict the earth in the form of two hemispheres, the parallels of latitude cease to be parallel lines. If we resort to what is called a rectangular projection, then the distances between latitudes cease to be equal. There is no more Euclidean geometry!

If you want to, I can show you that the Pythagorean theorem has failed. Take a map of important airlines (Figure 7.2) and take the triangle Moscow-Capetown-London. On the map it turns out to be an almost exact right triangle and so the sum of the squares on the sides must be equal to the square on the hypotenuse. Well, you can guess again. Let's figure it out exactly: the distance from Moscow to London is 2490 km, from Moscow to Capetown it is 10 130 km, and from London to Capetown 9660 km. The theorem fails completely. Our geometry does not hold true at all on a map. The laws of geometry in a plane depicting the globe differ from "ordinary" laws.

Take a look at the map of the hemispheres and you will see it has "edges". But this is an illusion. Actually, if one moves over the surface of the earth, he never comes to any "edge of the earth".

There is even a joke to the effect that Einstein's son asked his father why he was so famous. Said the great Einstein: "I was lucky, I was the first to notice that a bug crawling on the globe can crawl around the equator and come back to the starting point." That observation does not represent a discovery of course. But extend it to the three-dimensional space of the universe; maintain that the universe is finite and closed like a two-dimensional surface bounding the globe (thus laws of "ordinary" geometry do not hold true any longer); then from that draw the conclusion that all points of the universe are of an equal status in the same sense that all points on the surface of the globe are—well, all that requires exceptional intellectual courage.

Now we can conclude that if we earthlings observe that all galaxies are racing away from us in all directions, then an inhabitant of the planet of any star will observe the same; he will arrive at the same conclusions concerning the nature of motion of the stellar world and will measure the same velocities of the galaxies as the earth dweller.

The model of the universe advanced by Einstein in 1917 was a natural consequence of his so-called general theory of relativity (that part of the theory that we discussed in Chapter 4 is called the special theory of relativity).

However, Einstein did not envisage the possible expanding of a closed universe. The fact that a closed universe must expand was demonstrated in 1922-1924 by the Soviet scientist Alexandr A. Friedman (1888-1925). It turned out that the theory requires that the universe either expand or have alternating expansions and contractions. At any rate it cannot be static. We can accept either one of these two viewpoints, that is, we can assume that we are now living in an epoch of expansion of the universe, or we can assume that the universe was a "cosmic egg" some time ago (calculations show this time lapse to be equal to several tens of thousands of millions of years) and that it exploded and the debris has since been expanding.

It is important to have a clear understanding that the idea of an initial explosion is in no way connected with the concept of a creation of the world. It may very well be that attempts to look too far into the future and too far into the past and also out to excessively great distances are unjustified within the framework of existing theories.

Let us take a simple example and examine it in the light of a scheme that appears today to be quite reasonable. We measure the red shift of the spectral lines of radiation coming to us from distant galaxies. Using the Doppler equation, we estimate the velocities of the stars. The farther they are away from us, the faster they appear to be moving. Telescopes report speeds of recession of still more distant galaxies: ten thousand kilometres a second, a hundred thousand kilometres a second, and faster still. Now, there must be a limit to these recessional speeds. The point is that if a galaxy is moving away from us with the velocity of light, we will in principle never be able to see it, and all because, according to the Doppler equation, the frequency of light will vanish (become zero). No light would ever reach us from such a galaxy.

What are the greatest distances we are capable of measuring with the very best of the excellent instruments at our disposal? Our estimate will of course be extremely approximate. At any rate, we should not complain about not being able to look far enough away: the distances we are talking about run into thousands of millions of light years!

To speak of still greater distances is clearly meaningless. We might put it thus: within the framework of existing ideas, to talk about distances exceeding thousands of millions of light years is physically meaningless for the simple reason that no method has been devised for measuring such distances.

The situation here is much like that concerning the trajectory of an electron: it cannot be measured for the simple reason that the very concept is meaningless.

The General Theory of Relativity

The special theory of relativity made it necessary to introduce corrections into the laws of mechanics of bodies moving with velocities close to the velocity of light. The *general theory of relativity* introduces corrections into customary concepts concerning space when we deal with great distances. It is precisely for this reason that a discussion of the general theory is fitting in a chapter devoted to the physics of the universe.

The general theory of relativity rests on the following principle: there are no experiments capable of distinguishing the motion of bodies under the action of a gravitational field from the motion of such bodies in an appropriately chosen noninertial reference frame.

Let us consider a few simple examples. We are in a lift falling with acceleration a. Stretch out your hand and drop a ball. How will it fall? As soon as it is released, it will begin (according to the viewpoint of the inertial observer) its free fall with acceleration g. Since the lift is falling with acceleration a, the acceleration with respect to the floor of the lift will be $(g—a)$. An observer in the lift can describe the motion of the falling body with the help of the acceleration $g'=g—a$. In other words, an observer in the lift need not speak of any acceleration of the lift by "changing" the acceleration of the field of gravity in his frame of reference.

Now let us compare two lifts. One hangs motionless above the earth, the other is in motion in interplanetary space with acceleration a with respect to the stars. All bodies in the lift that is stationary above the earth can fall freely with acceleration g. And so can bodies inside the interplanetary lift. They will "fall", with acceleration $—a$, to the "bottom" of the lift. Here, the bottom will be the wall opposite the direction of acceleration.

It turns out that the action of a gravitational field and manifestations of accelerated motion are indistinguishable.

The behaviour of a body in a reference frame undergoing accelerated motion is the same as the behaviour of a body in the presence of an equivalent gravitational field. However, this equivalence can be complete if we confine ourselves to observations over small portions of space. Indeed, imagine a "lift" with a floor measuring thousands of kilometres. If such a lift hangs motionless above the earth, the phenomena taking place in it will

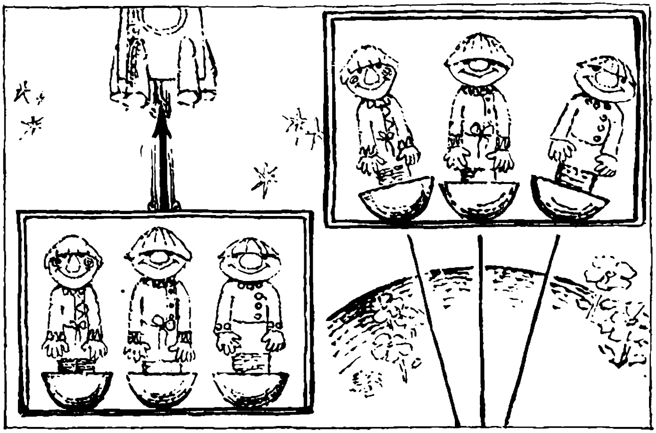


Figure 7.3

be different than in the case where the lift is moving with acceleration *a* relative to the fixed stars. This is clear from Figure 7.3: in one case the bodies fall at an angle to the bottom of the lift, in the other case they fall vertically.

Thus, to summarize, the principle of equivalence holds true for portions of space in which the field may be regarded as uniform.

The principle of equivalence of a gravitational field and a properly chosen local reference frame leads us to an important conclusion: a gravitational field is linked up with the curvature of space and a distortion in the flow of time.

Two observers are engaged in measuring distances and time intervals. They are interested in events occurring on a rotating disk. One observer is located on the disk, the other is motionless (relative to the stars). Incidentally, only the researcher on the disk is working. The fixed observer merely watches the work of his colleague.

The first experiment consists in measuring the radial distance, that is, the distance between two objects located on the same radius of the disk at different distances from the centre. The measurements are performed in the ordinary way: a standard ruler is laid off between the ends of the segment of interest. From the viewpoint of both investigators, the length of the ruler laid perpendicularly to the direction of motion is the same. And so there will be no argument between the two observers as to the length of the radial line-segment.

Now the disk dweller initiates his second experiment. He would like to measure the length of the circle. The ruler must now be laid down along the direction of motion, and of course we have to take into account the curvature of the circle. We therefore use a small ruler so that we can equate the length of a tangential line-segment with the length of the arc. The observers will not argue about the number of times the ruler fitted into the length of the circumference of the circle. And nevertheless their opinions about the length of the circumference will differ. The point is that the fixed observer will claim that the ruler contracted because, in the second experiment, it lay along the direction of motion.

And so the radius of the circle is the same for both observers but the length of the circumference of the circle is different. The fixed observer concludes that the formula for the length of the circumference of a circle, $2\pi r$, is incorrect. He says that the length of the circumference is greater than $2\pi r$.

This example shows us how the theory of relativity comes to a rejection of Euclidean geometry or (what is the same thing only said in different words) to the concept of curved space.

Similar "madness" is seen in the clock experiment.

Clocks fixed at different distances from the axis of rotation have different rates, and they are all slower than the fixed clock. Also, the farther a clock is from the centre of the disk the slower it goes. A fixed observer would say that you can use clocks and rulers on the disk only if you are located at a definite distance from the centre. Space and time possess local peculiarities.

Now let us recall the equivalence principle. Since local peculiarities of time and space manifest themselves on a rotating disk, this means that phenomena in a gravitational field behave in the same manner. The situation on the disk is the same as in the lift depicted in Figure 7.3. Accelerated motion is indistinguishable from gravitation acting against acceleration (in the direction opposite acceleration).

What this means is that a local curvature of space and time is equivalent to the presence of a gravitational field.

The closed nature of the universe that we discussed in the preceding section can undoubtedly be regarded as corroborating the general theory of relativity.

By applying rigorous mathematical reasoning, it is possible to derive a number of quantitative corollaries from the intricate equations of the general theory of relativity. First, Einstein demonstrated that a ray of light passing close to the sun must be deflected. The deflection of a ray of light passing very close to the sun should amount to 1.75 seconds of arc. Actual measurements yielded a figure of 1.70. Second, the orbit of the planet Mercury (the perihelion of the orbit, to be precise) should turn in its plane. Calculations show that in one century this should amount to 43 seconds of arc. Experiment has revealed precisely that number. And now a third prediction that was confirmed by experiment: a photon used up energy (and hence the frequency of the light changes) in overcoming gravitational forces.

The general theory of relativity is one of the greatest attainments of human thought. It has played a tremendous role in the development of our ideas about the universe and has revolutionized physics.

Stars of All Ages

The physics of the universe is still growing fast, it cannot at all be regarded as a fully developed science like, say, the mechanics of low velocities or thermodynamics. There is therefore every reason to believe that stellar investigations will lead to discoveries of new laws of nature. So far this has not occurred, but the picture of the universe drawn from time to time by the popularizing physicist is constantly changing. What I have to say here and now will, in a decade or so, most likely undergo considerable change.

Astronomers have long realized that there are different kinds of stars. With the help of the telescope, the spectrograph, and the interferometer we can determine many physical quantities that go to classify the stars.

As may be expected by analogy with terrestrial experiments (see page 12), the maximum intensity of the spectrum defines the surface temperature of a star. The observed colour of the star is uniquely related to this temperature. If the temperature is from 3 to 4 thousand degrees, the colour is reddish, if the temperature is between 6 and 7 thousand degrees, the star is yellowish. Pale blue stars have temperatures exceeding 10-12 thousand degrees. When physicists were able to get out into space, they found stars whose maximum radiation lies in the region of X rays and gamma rays. This means that the temperatures of such stars can reach into millions of degrees.

Another important characteristic of a star is the total energy of the spectrum that reaches us. This is the *luminosity* of the star. The colossal differences in luminosity may be connected with the size of the star, its distance from us, and its temperature.

As to chemical composition, stars consist largely of hydrogen-helium plasma. Our sun is a rather typical star. Its chemical composition has been determined more or less precisely from the type of spectrum and from theoretical calculations of the energy of its radiation. Hydrogen makes up 82% and helium 18%. All other elements come to only 0.1% of the total mass of the sun.

The atmospheres of many stars exhibit powerful magnetic fields that are thousands of times stronger than the magnetic field of the earth. We know about this due to spectral analysis, since the spectral lines split in magnetic fields.

The interstellar medium is rarefied to an extent beyond all imagination: one cubic centimetre of outer space contains one atom. This is a good time to recall that one cubic centimetre of the air we breathe contains 2.7×10^{19} molecules, on the average. There are regions of space where the density of the interstellar gas is appreciably higher than the average. Besides gas we encounter dust, which consists of particles of size 10^{-4} to 10^{-5} cm.

It can be assumed that stars are formed out of this gaseous-dust medium. Under the influence of gravitational forces, a cloud begins to contract into a sphere. In the course of hundreds of thousands of years it contracts and the temperature of the star increases making it visible in the heavens. Of course, the time required for this depends on the size and, hence, the mass of the condensing cloud.

As the contraction continues, the temperature in the interior of the star rises and attains a value at which a thermonuclear (fusion) reaction can start up. Four nuclei of the hydrogen atom combine into a single nucleus of a helium atom. Recall that in the process, 4.0339 atomic mass units of four hydrogen atoms are converted into

4.0038 atomic mass units of the helium atom. Thus a mass equal to 0.0301 unit is converted into energy.

The burning of hydrogen that takes place in the deep interior of a star can continue for different time intervals depending on the mass of the star. For the sun, this time interval is equal to between 10 and 20 thousand million years. That is the stable-state period of the star. The forces of gravitational attraction are balanced by the internal pressure of the hot nuclei that tend to blow up the star. A star is hence something like a gas cylinder. The walls of the cylinder, so to speak, are the gravitational forces.

When the supply of hydrogen fuel comes to an end, the internal pressure diminishes, and the core of the star begins to collapse.

What happens next? To that question the theoretician replies that everything depends on whether the star is able to shrug off its outer shell. If the star does throw off the outer shell, the mass of the star becomes roughly half that of the sun, and then forces take over that are capable of counteracting the gravitational forces. The result is a tiny star with a high surface temperature. It goes by the name of *white dwarf*.

What next? Again the fate of the star depends on the mass. If the white dwarf has a mass less than one and a half solar masses, then it will die out slowly with no dramatic events taking place. The radius will diminish and the temperature will fall. Ultimately, the dwarf will turn into a cool star the size of the earth. Such is the demise of most stars.

Now if the mass of the white dwarf that took shape after the star with fuel expended has thrown off its outer shell is more than one and a half solar masses, then contraction does not come to a halt at the white-dwarf stage. Electrons merge with protons to form a *neutron star*, which measures only a few tens of kilometres across.

Calculations show that a neutron star should have a temperature of the order about ten million degrees. It has a maximum radiation in the region of X rays.

We have just discussed the history of a star that is able to throw off the outer shell. However, the mathematical equations do not insist on such disrobing. Now if a celestial body retains the mass of, say, ten of our suns, then the gravitational attraction would simply destroy the star. Where the star once was there would be a *black hole*.

At what stage in the compression process should this take place, and what does the term "black hole" stand for?

Let us recall the simple laws that underlie the launching of space craft from the earth (see the first book in this series). To escape from the earth's gravitational pull requires a velocity of 11 kilometres per second. This velocity is given by the equation

$$v^2 = \gamma \frac{M}{R}$$

From this equation it is evident that as a sphere of definite mass contracts, the escape velocity of a rocket from such a body into outer space will be constantly increasing. But the limiting velocity is equal to 300 000 kilometres per second! If a stellar sphere of given mass contracts to a ball having radius

$$R = \gamma \frac{M}{(300\,000\ \text{km/s})^2}$$

then it becomes impossible to escape. In other words, anything at all can go into what was once a star (say, a ray of light or other electromagnetic radiation), but it cannot go out; it cannot escape from the hole. Black hole is indeed a very appropriate designation. Using the above equation, we can see at once that black holes with from 3 to 50 solar masses will be from 60 to 1000 kilometres across.

A few words are now in order about how to search for black holes. The reader may think that this problem is too small for a little book devoted to the whole of physics, but I think the very method of approaching such a search is highly instructive. The talent of a scientist is revealed precisely in the search for indirect proofs of a model whose properties cannot be demonstrated directly.

At first glance, the problem does indeed appear to be exceedingly complicated, if not downright unsolvable. A black spot in the sky 1000 kilometres across corresponds to a millionth of one second of arc. No telescope could detect a black hole.

Over twenty years ago the Soviet physicist Ya. Zeldovich proposed a search for black holes based on the idea that their presence in the sky should affect the behaviour of nearby visible bodies. Together with his co-workers, he conducted a systematic examination of star catalogues in order to find a visible star that might be revolving about a black hole. Such a star would appear to be alone but its revolution would indicate that the spectral lines would periodically be displaced redwards or bluewards depending on whether the star was moving away from us or towards us.

Research workers in other countries also entered the search and a certain number of suitable stars were found. From the amount of the Doppler shift we can give a rough estimate of the mass of the star about which the visible satellite star is revolving. Some invisible candidates were selected with masses three times the solar mass. Thus, these could not be white dwarfs or neutron stars.

And yet all of this is not enough to claim that such an exotic entity as a black hole does indeed exist. Any opponent could advance a whole series of explanations for the periodic Doppler shift.

However, there is one phenomenon that we can call to our aid. A black hole is capable of drawing into itself gas from its satellite star. As this gas falls into the black hole, it should heat up intensely and should emit X rays. True, neutron stars and white dwarfs are also capable of drawing in gas, but such stars, as we have already pointed out, can be distinguished from a black hole by their mass.

Just recently a star was found that satisfied all requirements for the satellite of a black hole. New experiments will undoubtedly follow and also detailed theoretical calculations whose aim is to predict the peculiarities of the X-ray spectrum emanating from the surrounding space of a black hole. The near future will surely show us how often these amazing "bodies" turn up in the universe. There are grounds to believe that there may be large black holes and mini-holes of mass of the order of 10^{-16} gram. Such holes less than an atomic nucleus in size may suddenly die out and in their final act return the energy contained in them. This energy would be so great as to satisfy all terrestrial needs for many years! What a marvelous topic for science fiction writers.

Radio Astronomy

The photograph shown in Figure 7.4 depicts a parabolic radio antenna that focuses incident parallel radio beams. The beams come to a point and enter a special receiver. The signal is then amplified by electronic equipment. The parabolic antenna shown in the figure is that of the Effelsberg Radio Observatory (near Bonn, West Germany). This fully steerable radio telescope (about 100 metres in diameter) is used in joint research by teams of scientists from many countries, including the USSR.

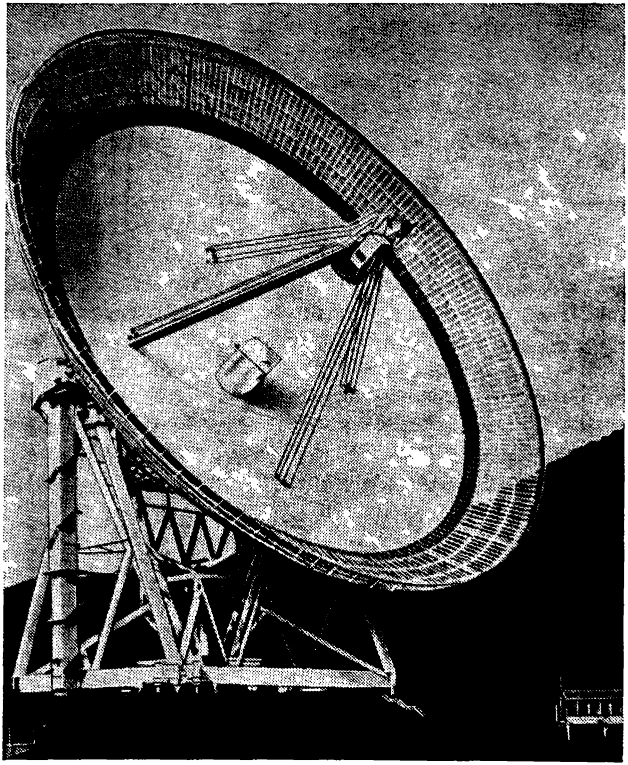


Figure 7.4

Such telescopes are extremely sensitive. They can be steered in any direction and the antenna is capable of detecting energy fluxes of the order of 10^{-28} watt second per square metre. Fantastic!

Radio astronomy has led to fundamental discoveries in the physics of the universe.

In the not too distant future radio telescopes will be set up on the moon and on artificial earth satellites. Then the absorption and reflection of electromagnetic waves by the earth's atmosphere will cease to be an obstacle to the observer. So far there are two "windows" in the electromagnetic spectrum. One of them lets in visible light, the other radio waves between 2 centimetres (15 000 megahertz) and 30 metres (10 megahertz).

The weather has no effect on radio-astronomy observations. The radio sky is quite different from what we see at night. Powerful radio stars, including many galaxies and the so-called *quasars* (or *quasi-stellar objects*), are hardly at all visible in the light spectrum.

The radio emission of outer space is not very strong and its study became possible only due to the phenomenal achievements of radio electronics. Suffice it to say that the radio emission of the sun is a million times less powerful than the emission in the optical portion of the spectrum.

And yet, without radiospectroscopy we would not have been able to establish many important facts. For instance, a big role in understanding processes occurring in the universe is played by measurements of the residual emission of explosions of *supernovae*.

Neutral hydrogen emits a strong wave of 21 centimetres. Measurements of the intensity of this radio emission have enabled scientists to sketch a pattern of the distribution of interstellar gas in space and follow the movements of gaseous clouds.

A large number of radio galaxies and quasars have been found at distances extending to the very limit of observational capabilities. Suffice it to say that the red shift in the radiation coming from these sources reaches a value of 3.5. The *red shift* is defined as the ratio of the difference between the emitted and received wavelength to the magnitude of the emitted wavelength. This means the difference is 3.5 times greater than the wavelength of the radiation.

Radio methods have enabled us to look out to the very fringes of the universe. Radio-astronomy investigations have enabled us to clarify the nature of the cosmic radiation that comes to us from the outer reaches of the universe.

Cosmic Rays

Studies that are now conveniently conducted in outer space beyond the earth's atmosphere demonstrate that our planet is under a constant bombardment by streams of nuclear particles moving at velocities close to the velocity of light itself. These particles have energies between 10^8 and 10^{20} electron volts. An energy of the order of 10^{20} electron volts is eight orders of magnitude greater than the energy of the most powerful particle accelerators we have been able to build.

The primary cosmic radiation consists mainly of protons (about 90%); the remainder is made up of heavier nuclei. Naturally, when cosmic-ray particles collide with molecules, atoms, and nuclei, they can make elementary particles of all types. But astrophysicists are particularly interested in the primary radiation. How are such energetic particle fluxes created? Where are the sources of cosmic rays located?

A long time ago it was demonstrated that the sun is not the main source of cosmic radiation. But if that is so, then the responsibility for generating cosmic rays cannot be shifted to other stars either, for they are in no way different (in principle) from our sun. Then who is to blame?

In our Galaxy we have what is known as the Crab Nebula that was formed in the explosion of a star in the year 1054 (don't forget that the sky has been under observation for many thousands of years). This has been shown to be a source of radio waves and cosmic radiation—a coincidence that resolves the mystery of the tremendous energies of cosmic protons. As we know, a particle can build up energy by spiralling about a magnetic line of force. All we need to do is presume that the electromagnetic field generated in the explosion of the star plays the role of a synchrotron, and then the enormous energy acquired by a particle as it journeys along a spiral curve about a line of force over periods of thousands of years can reach truly fantastic figures (like those we have mentioned).

Calculations show that after covering a distance equal to the diameter of our Galaxy, a cosmic particle cannot acquire more energy than 10^{19} electron volts. Apparently, particles of maximum energy come to us from other galaxies.

Naturally we need not assume that only stellar explosions generate cosmic radiation. Any sources of radio waves can also be sources of cosmic radiation.

Cosmic rays were discovered at the start of this century. Electroscopes carried aloft in balloons revealed some exciting facts. Researchers found that the discharge of an electroscope at high altitudes proceeds much more quickly than it does at sea level.

It became clear that the usual falling of the leaves of the electroscope is not the result of imperfections in the instrument, but a phenomenon due to the action of some kind of external factors.

In the 1920s, physicists were quite convinced that the ionization of the air that removes the charge from an electroscope is undoubtedly of extraterrestrial origin. The American physicist Robert Andrews Millikan (1868-1953) was the first to state this confidently and he gave the phenomenon its modern name: *cosmic rays* (or *cosmic radiation*).

In 1927, the Soviet scientist D. V Skobeltsyn obtained the first photograph of the tracks of cosmic rays in an ionization chamber.

The energies of cosmic particles were determined by methods that we have already described elsewhere. They turned out to be tremendous.

In their studies of cosmic rays, physicists have made a number of very remarkable discoveries. For one thing, the existence of the positron was demonstrated in precisely this way. The same goes for *mesons*, which are particles with mass intermediate between the mass of the proton and that of the electron; they too were first detected in cosmic radiation.

Studying cosmic rays continues to be one of the most exciting fields of physical investigation.

* * *

The very incompleteness of astrophysics makes it difficult to describe it in a single chapter of a small book devoted to merely an introduction of the reader to the basic facts and ideas of physics at large. I could only choose a few of the many physical problems that deal with the universe, only those that I consider to be of special interest.

To the Reader

Mir Publishers would be grateful for your comments on the content, translation and design of this book. We would also be pleased to receive any other suggestions you may wish to make.

Our address is
Mir Publishers
2 Pervy Rizhsky Pereulok
I-110, GSP, Moscow, 129820
USSR

Printed in the Union of Soviet Socialist Republics

OTHER BOOKS IN THE SERIES

BOOK 1 **PHYSICAL BODIES**

BOOK 2 **MOLECULES**

BOOK 3 **ELECTRONS**

The aim of the authors, the late Lev Landau, Nobel and Lenin prize winner, and Prof. Alexander Kitaigorodsky, was to provide a clear understanding of the fundamental ideas and latest discoveries in physics, set forth in readily understandable language for the laymen. The new and original manner in which the material is presented enables this book to be enjoyed by a wide range of readers, from those making the first acquaintance with physics to college graduates and others interested in this branch of science. These books may serve as an excellent instructional aid for physics teachers.

Physical Bodies

L. D. LANDAU, A. I. KITAIGORODSKY

In this first book of the famous series *Physics for Everyone* by the late renowned scientist, winner of the Nobel and Lenin prizes, Academician Lev Landau, and the distinguished physicist Alexander Kitaigorodsky, the motion of bodies is dealt with from the points of view of two observers: one in an inertial and the other in a non-inertial frame of reference. The law of universal gravitation is discussed in detail, including its application for calculating space velocities and for interpreting lunar tides and various geophysical phenomena.
The principal aim of the book is to explain to a layman reader in an easy-to-understand manner the basic concepts and latest developments in the field of physics.