

* *
* *

“Où en êtes-vous? Que faites-vous? Il faut travailler” (on his death-bed, to his devoted pupils, watching over him).

The Spontaneous Generation Controversy **(340 BCE–1870 CE)**

“Omne vivium ex Vivo.”

(Latin proverb)

Although the theory of spontaneous generation (*abiogenesis*) can be traced back at least to the Ionian school (600 B.C.), it was Aristotle (384–322 B.C.) who presented the most complete arguments for and the clearest statement of this theory. In his “On the Origin of Animals”, **Aristotle** states not only that animals originate from other similar animals, but also that *living things do arise and always have arisen from lifeless matter*. Aristotle’s theory of spontaneous generation was adopted by the Romans and Neo-Platonic philosophers and, through them, by the early fathers of the Christian Church. With only minor modifications, these philosophers’ ideas on the origin of life, supported by the full force of Christian dogma, dominated the mind of mankind for more than 2000 years.

According to this theory, a great variety of organisms could arise from lifeless matter. For example, worms, fireflies, and other insects arose from morning dew or from decaying slime and manure, and earthworms originated from soil, rainwater, and humus. Even higher forms of life could originate spontaneously according to Aristotle. Eels and other kinds of fish came from the wet ooze, sand, slime, and rotting seaweed; frogs and salamanders came from slime.

Rather than examining the claims of spontaneous generation more closely, Aristotle's followers concerned themselves with the production of even more remarkable recipes. Probably the most famous of these was **van Helmont's** (1577-1644) recipe for mice. By placing a dirty shirt into a bin containing wheat germ and allowing it to stand 21 days, live mice could be obtained. Another example was the slightly more complicated but equally "foolproof" recipe for bees. By killing a young bullock with a knock on the head, burying him in a standing position with his horns sticking out of the ground, and finally sawing off his horns one month later, out will fly a swarm of bees.

The more exact methods of observation that were developed during the seventeenth century soon led to a realization of the complex nature of the anatomy and life cycles of certain living organisms. Equipped with this better understanding of the complexity of living organisms, it became more difficult for some to accept the theory of spontaneous generation. This skepticism signaled the beginning of three centuries of heated controversy over a theory that had gone unchallenged for the previous 2000 years. What is more significant is that the controversy was to be resolved not by powerful arguments but by ingeniously designed, simple experiments.

The controversy went through four phases:

I. REDI (1688) VS. ARISTOTELIAN SCHOOL AND CHURCH DOGMA

Redi was first to use carefully controlled experiments to test the theory of spontaneous generation. He put some meat in two jars. One he left open to air (the control); the other he covered securely with gauze. At that time it was well recognized that white worms would arise from decaying meat or fish. Sure enough, in a few weeks, the meat was infested with the white worms but only in the control jar which was not covered. This experiment was repeated several times, using either meat or fish, with the same result. On closer examination he noted that common houseflies went down into the meat in the open jar, later the white worms appeared, and then new flies. Redi reported that he had observed the flies deposit their eggs on the gauze; however, worms developed in the meat only when the eggs got to the meat. He therefore concluded from his observations that the white worms did not arise from the putrid meat. The worms developed from the eggs that the flies had deposited. The white worm then was the larva of the fly, and the meat served only as food for the developing insect.

Redi's experiment provided the impetus for testing other well-established recipes. In all cases that were examined carefully, it was demonstrated that the living organism arose not by spontaneous generation, but from a parent. Thus it was shown that the theory of spontaneous generation was based on a combination of the weakness of the human eye and bits and snatches of information gathered by accidental observation. The early biologists had seen earthworms coming out of the soil and frogs emerging from the slime of pond water, but they had not been able to see the tiny eggs from which these organisms arose. Because their observations had not been systematic, they had not seen how the mice invaded the grain bin in search of food, so they thought that the grain produced the mice. Based on the more exact methods of observation, the evidence that supported the theory of spontaneous generation of animals and plants was largely demolished by the end of the seventeenth century.

II. SPALLANZANI VS. NEEDHAM (1767–1768)

*As soon as the discoveries of Leeuwenhoek¹⁹⁹ became known, the proponents of spontaneous generation turned their attention to these microscopic organisms and suggested that surely they must have formed by spontaneous generation. Finally, experimental “proof” for this notion was published in 1748 by an Irish priest, **John Tuberville Needham** (1713–1781).*

*Needham reported that he had taken mutton gravy fresh from the fire, transferred it to a flask, heated it to boiling, stoppered it tightly with a cork, and then set it aside. Despite boiling, the liquid became turbid in a few days. When examined under a microscope, it was teeming with microorganisms of all types. The experiments were repeated by and gained the support of the famous French naturalist, **Georges Louis Le-clerc, Comte de Buffon** (1707–1788). Needham's demonstration of spontaneous generation was generally accepted as a great scientific achievement, and he was immediately*

¹⁹⁹ The development of *microscopy* started with **Janssen** (1590) and continued with **Hooke** (1660), **Leeuwenhoek** (1676) and **Zeiss** (1883). Just as the theory of the abiogenesis of higher organisms was being refuted, the controversy was reopened, more heated than ever, because of the discovery of microorganisms by Antony van Leeuwenhoek. Leeuwenhoek patiently improve his microscopes and developed his techniques of observation for 20 years before he reported any of his results.

elected into the Royal Society of England and the Academy of Sciences of Paris.

Meanwhile in Italy, Lazzaro Spallanzani (1729-1799) performed a series of brilliantly designed experiments of his own that refuted Needham's conclusions. Spallanzani found that if he boiled the food for one hour and hermetically sealed the flasks (by fusing the glass so that no gas could enter or escape), no microorganisms would appear in the flasks. If, however, he boiled the food for only a few minutes, or if he closed the flask with a cork, he obtained the same results that Needham reported. Thus he wrote that Needham's conclusions were invalid because (1) he had not heated the gravy hot enough or long enough to kill the microorganisms, and (2) he had not closed the flask sufficiently to prevent other microbes from entering.

Count Buffon and Father Needham immediately responded that, of course, Spallanzani did not generate microorganisms in his flasks because his extreme heating procedures destroyed the *vegetative force* in the food and the *elasticity* of the air. Regarding Spallanzani's experiments, Needham wrote, "from the way he has treated and tortured his vegetable infusions, it is obvious that he has not only much weakened, and maybe even destroyed, the vegetative force of the infused substances, but also that he has completely degraded ... the small amount of air which was left in his vials. It is not surprising, thus, that his infusions did not show any sign of life."

Rather than engage in theoretical arguments over the possible existence of these mystical forces, Spallanzani returned to the laboratory and performed another set of ingenious experiments. This time he heated the sealed flasks to boiling not for one hour, but for three hours, and even longer. If Needham was right, this treatment should certainly have destroyed the vegetative force. As Spallanzani had previously observed, nothing grew in these heated, sealed flasks. However, when the seal was broken and replaced with a cork, the broth soon became turbid with microbes. Since even three hours of boiling did not destroy anything in the food necessary for the production of microbes, Needham could no longer argue that he had killed the vegetative force by the heat treatment.

Spallanzani continued to perform experiments that led him to the conclusion that properly heated and hermetically sealed flasks containing broth would remain permanently lifeless. He was, however, unable to answer adequately the criticism that by sealing the flasks he had excluded the "vital forces" in the air that Needham claimed were also necessary ingredients for spontaneous generation. With the discovery of oxygen gas in 1774 and the realization that this gas is essential for the growth of most organisms, the possibility that spontaneous generation could occur, but only in the presence of air (oxygen), gained additional support.

III-1. SCHWANN VS. BERZELIUS, LIEBIG AND WHOLER (1836–1839)
 – THE FERMENTATION CONTROVERSY

The art of brewing was developed by trial and error over a 6000-year period and practiced without any understanding of the underlying principles. From long experience, the brewer learned the conditions, not the reasons, for success. Only with the advent of experimental science in the eighteenth and nineteenth centuries did man attempt to explain the mysteries of fermentation. Let us, then, from our vantage point in time, trace the observations, experiments, and debates from which evolved our present understanding of fermentation and biological catalysis.

For centuries, fermentation had a significance that was almost equivalent to what we would now call a chemical reaction, an error that probably arose from the vigorous bubbling seen during the process. The conviction that fermentation was strictly a chemical event gained further support during the early part of the nineteenth century, when French chemists led by **Lavoisier** and **Gay-Lussac** determined that the alcoholic fermentation process could be expressed chemically by the following equation:



It was, of course, known that yeast must be added to the wort in order to ensure a reproducible and rapid fermentation. The function of the yeast, according to the chemists, was merely to catalyze the process. All chemists agreed that fermentation was in principle no different from other catalyzed chemical reactions.

Then in 1837, the French physicist **Charles Cagniard-Latour** and the German physiologist **Theodor Schwann** independently published studies that indicated yeast was a living microorganism. Prior to their publications, yeast was considered a proteinaceous chemical substance. The reason the two workers came up with the same observations at approximately the same time is most likely due to the production of better microscopes.

It should be mentioned that one of the reasons it was difficult to ascertain whether or not yeast is living was because, like most other fungi, yeast is not motile. The organized cellular nature of yeast was discovered only when improved microscopes became available. Schwann and Cagniard-Latour also observed that alcoholic fermentation always began with the first appearance of yeast, progressed only with its multiplication, and ceased as soon as its growth stopped. Both scientists concluded that alcohol is a by-product of the growth process of yeast.

The biological theory of fermentation advanced by Cagniard-Latour and Schwann was immediately attacked by the leading chemists of the time. The

eminent Swedish physical chemist **Jons Jakob Berzelius** reviewed the two papers in his *Jahresbericht* for 1839 and concluded that microscopic evidence was of no value in what was obviously a purely chemical problem. According to Berzelius, nothing was living in yeast.

This opinion was supported by **Justus von Liebig** and **Friedrich Wöhler**. Liebig argued that:

1. Certain types of fermentation, such as the lactic acid (souring of milk) and acetic acid (formation of vinegar) fermentations, can occur in the complete absence of yeast.
2. Even if yeast is living, it is not necessary to conclude that the alcoholic fermentation is a biological process. The yeast is a remarkably unstable substance which, as a consequence of its own death and decomposition, catalyzes the splitting of sugar. Thus, fermentation is essentially a chemical change catalyzed by breakdown products of the yeast.

Liebig's views were widely accepted, partly because of his powerful influence in the scientific world and partly because of a desire to avoid seeing an important chemical change relegated to the domain of biology. And so the stage was set – biology against chemistry – for the entrance of Louis Pasteur.

III-2. PASTEUR VS. LIEBIG AND BERZELIUS (1857–1860)

In 1851, **Pasteur** published his first paper on the topic of fermentation. The publication dealt with lactic acid fermentation, not alcoholic fermentation. Utilizing the finest microscopes of the time, Pasteur discovered that souring of milk was correlated with the growth of a microorganism, but one considerably smaller than the beer yeast. During the next few years, Pasteur extended these studies to other fermentative processes, such as the formation of butyric acid as butter turns rancid. In each case he was able to demonstrate the involvement of a specific and characteristic microorganism; alcoholic fermentation was always accompanied by yeasts, lactic acid fermentation by nonmotile bacteria, and butyric acid fermentation by motile rod-shaped bacteria. Thus, Pasteur not only disposed of one of the opposition's strongest arguments, but also provided powerful circumstantial evidence for the biological theory of fermentation.

Now Pasteur was ready to attack the crucial problem, alcoholic fermentation. Liebig had argued that this fermentation was the result of the decay of

yeast; the proteinaceous material that is released during this decomposition catalyzes the splitting of sugar. Pasteur countered this argument by developing a protein-free medium for the growth of yeast. He found that yeast could grow in a medium composed of glucose, ammonium salts, and some incinerated yeast. If this medium is kept sterile, neither growth nor fermentation takes place. As soon as the medium is inoculated with even a trace of yeast, growth commences and fermentation ensues. The quantity of alcohol produced parallels the multiplication of the yeast. In this protein-free medium, Pasteur was able to show that fermentation takes place without the decomposition of yeast. In fact, the yeast synthesizes protein at the expense of the sugar and ammonium salts. Thus Pasteur concluded in 1860:

“Fermentation is a biological process, and it is the subvisible organisms which cause the changes in the fermentation process. What’s more, there are different kinds of microbes for each kind of fermentation. I am of the opinion that alcoholic fermentation never occurs without simultaneous organization, development and multiplication of cells, or continued life of the cells already formed. The results expressed in this memoir seem to me to be completely opposed to the opinion of Liebig and Berzelius.”

Pasteur argued effectively, and more important, all the data were on his side. Thus the vitalistic theory of fermentation predominated until 1897, when an accidental discovery by **Eduard Buchner** (1860–1917) demonstrated that the alcoholic fermentation of sugars is due to action of *enzymes* contained in the yeast.

The controversy was thus finally resolved and the door was thrown open to modern biochemistry.

IV. PASTEUR AND TYNDALL VS. POUCHET (1859–1885)

The spontaneous generation controversy was brought to a crisis in 1859 when **Felix Archimède Pouchet** (1800-1872), a distinguished scientist and director of the Museum of Natural History in Rouen, France, reported his experiments on spontaneous generation. Pouchet claimed to have accomplished spontaneous generation using hermetically sealed flasks and pure oxygen gas. These experiments, he argued, demonstrated that “animals and plants could be generated in a medium absolutely free from atmospheric air and in which therefore no germ of organic bodies could have been brought by air.”

The impact of Pouchet’s experiments on his contemporaries was so great that the French Academy of Sciences offered the Alhumpert Prize in 1860 for

exact and convincing experiments that would end this controversy once and for all. Pasteur first set out to demonstrate that air could contain numerous microorganisms. From his microscopic observation, Pasteur concluded that there are large numbers of organized bodies suspended in the atmosphere. Furthermore, some of these organized bodies are indistinguishable by shape, size, and structure from microorganisms found in contaminated broths. Later he showed that these organized bodies that collected on the cotton fibers not only looked like microorganisms, but when placed in a sterile broth were capable of growth!

Pasteur's second series of experiments provided further circumstantial evidence that it was the microbes on floating dust particles and not the so-called vital forces that were responsible for sterilized broth's becoming contaminated. In these experiments, Pasteur carried sterile-sealed flasks to a wide variety of locations in France. At the various sites, he would break the seal, allowing air to enter the flask. The flask was immediately resealed and brought back to Paris for incubation. The conclusion from these numerous experiments was that where considerable dust existed, all the flasks would become turbid. For example, if he opened sterile flasks in the city, even for a brief period, they all became turbid, whereas in the mountainous regions, especially at high altitudes, a large proportion of the flasks remained sterile.

His third and most conclusive experiment utilized the now famous swan-neck flask. As a result of the experiments described, Pasteur hypothesized that the source of contamination was dust. If true, then it should be possible to keep a broth sterile even in the presence of air as long as the dust is kept out. In order to test this hypothesis, Pasteur constructed several bent-neck flasks. After placing broth into the flask, he boiled the liquid for a few minutes, driving the air from the orifice of the flask. As the flask cooled, fresh air entered the flask. Despite the fact that the broth was in contact with the gases of the air, the fluid in the swan-neck flask always remained sterile. Pasteur reasoned correctly that the dust particles that entered the flask were absorbed onto the walls of the neck and never penetrated into the liquid. As an experimental control, Pasteur demonstrated that nothing was wrong with the broth. If he broke the neck off the flask or tipped liquid into the neck (in both cases dust would enter the broth), the fluid soon became turbid with microorganisms.

With these simple, ingenious experiments, Pasteur not only overcame the criticism that air was necessary for spontaneous generation but he was also able to explain satisfactorily many of the sources (dust) of the contradictory findings of other investigators. Although Pasteur's conclusions gained wide support in both the scientific and the lay communities, they did not convince all the proponents of spontaneous generation.

Pouchet and his followers continued to publish reports of spontaneous generation. They claimed their techniques were as rigorous as those of Pasteur. Where Pasteur failed to obtain spontaneous generation they succeeded in every case. For example, they carefully opened 100 flasks at the edge of the Maladetta Glacier in the Pyrenees Mountains at an elevation of 10,850 feet. In this region which Pasteur had found to be almost dust free, all 100 of Pouchet's flasks became turbid after a brief exposure to the air. Even when Pouchet used swan-neck flasks, there was growth.

To Pasteur, this disagreement no longer revolved around the interpretation of experiments; rather, either Pouchet was lying or his techniques were faulty, Pasteur had complete faith in his own procedures and results and had no respect for those of his opponents. Thus he challenged Pouchet to a contest in which both of them would repeat their experiments in front of their esteemed colleagues of the Academy of Science. Pouchet accepted the challenge with the added statement, "If a single one of our flasks remains unaltered, we shall loyally acknowledge our defeat."

*A date was set, and the place was to be the laboratory in the Museum of Natural History at the Jardin des Plantes, Paris²⁰⁰. Pasteur arrived early with the necessary apparatus for demonstrating his techniques. Newspaper photographers and reporters were also on hand for this event of great public interest. However, Pouchet did not show up, and Pasteur won by default. It is difficult to ascertain whether Pouchet was intimidated by Pasteur's confidence or, as he later stated, he refused to take part in the "circus" atmosphere that Pasteur had created, and that their scientific findings should instead be reported in the reputable scientific journals. At any rate, in Pouchet's absence, Pasteur repeated his experiments in front of the referees with the same results he had previously obtained. As far as the scientific community was concerned, the matter was settled²⁰¹. The law *Omne vivium ex vivo* (All life from life) also applied to microorganisms.*

In retrospect, however, the most ironic aspect of this extraordinary contest was not that Pouchet failed to show up, but rather that if he had appeared, he would have won! Pouchet's experiments are reproducible. Pouchet performed his experiments in the following manner: He filled swan-neck flasks with a

²⁰⁰ **Henri Milne-Edwards** (1800–1885), a French naturalist and zoologist (then a professor at the Museum and from 1864, its director) lent political and scientific support to Pasteur during the Pasteur-Pouchet debate. He wrote important works on crustaceans, mollusks, and corals and wrote a major opus on comparative anatomy and physiology.

²⁰¹ Yet, the Pasteur-Pouchet debate had a chilling effect on French evolutionary research for decades.

broth made from hay, boiled them for one hour, and then allowed the flasks to cool. He obtained growth in every flask. Pasteur's experiments differed in only two respects. Pasteur used a mixture of sugar and yeast extract for broth and boiled it for just a few minutes. Pasteur never obtained growth in his swan-neck flasks. The reason for their contradictory results was not understood until 1877, 17 years later.

Mainly because of the careful work of the English physicist **Tyndall** (1820-1893), Pouchet's experiments could be explained without invoking spontaneous generation. Tyndall found that foods vary considerably in the length of boiling time required to sterilize them. For example, the yeast extract and sugar broth of Pasteur could be sterilized with just a few minutes of boiling, whereas the hay medium of Pouchet required heating for several hours to accomplish sterilization. Tyndall postulated that certain microorganisms can exist in heat-resistant forms, which are now referred to as spores. Furthermore, studies by Tyndall and the French bacteriologist **Ferdinand Cohen** revealed that hay media contain a large number of such spores. Thus the contradictory results of Pasteur and Pouchet were due to differences in the broths they used.

Tyndall went on to demonstrate that nutrient medium containing spores can be sterilized easily by boiling for one-half hour on three successive days. This procedure of discontinuous heating, now called *Tyndallization*, works as follows: The first heating kills all the cells that are not spores and induces the spores to germinate (in the process of germination, the spores lose their heat resistance as they begin to grow); on the second day, the spores have germinated and are thus susceptible to the heating. The third day heating "catches" any late germinating spores. Thus, with the publication of Tyndall's work, all the evidence that supported the theory of spontaneous generation was destroyed. Since that time, there has been no serious attempt to revive this theory.

It should be pointed out, however, that by its very nature, the *theory of spontaneous generation cannot be disproved*. One can always argue that the conditions necessary for spontaneous generation have not yet been discovered. Pasteur was well aware of the difficulty of a negative proof, and in his concluding remarks on the controversy, he merely showed that spontaneous generation had never been demonstrated.

There is no known circumstance in which it can be affirmed that microscopic beings came into the world without germs, without parents similar to themselves. Those who affirm it have been duped by illusions, by ill-conducted experiments, and by errors that they either did not perceive, or did not know how to avoid.

1847 CE Augustus De Morgan (1806²⁰²–1871, England). Mathematician and logician, a contemporary of **Boole**. Laid the foundation of modern *symbolic logic* and developed new technology for logical expressions. Formulated *De Morgan's laws*. Introduced and vigorously defined the term *mathematical induction*. He endeavored to reconcile mathematics and logic, but compared with Boole, his impact on modern mathematics and its applications is small²⁰³, and he is remembered mainly as a logical reformer. He is most noteworthy as the founder of the *logic of relations* and as a developer of the *algebra of logic* which reconstructed the logic of Aristotle upon a mathematical basis.

De Morgan was born in India, and taught at University College in London during 1836–1866. Although a convinced theist, he never joined a religious congregation. He renounced his professorship in 1866 when a colleague was denied a chair at University College because he was a unitarian.

The Basic Ideas of Topology

I. POLYHEDRA AND SURFACES²⁰⁴

A simple polyhedron is a body enclosed by faces, all of which are plane polygons (some examples of polyhedra are: pyramid, prism, frustum). It has

²⁰² De Morgan was always interested in odd numerical facts; thus in 1849, he noticed that he had the distinction of being x years old in the year x^2 ($x = 43$).

²⁰³ Nevertheless, he shall be remembered in mathematics proper due to his discovery of the summation formula:

$$\sum_{n=1}^N \frac{x^{2^{n-1}}}{x^{2^n} - 1} = \frac{1}{x - 1} - \frac{1}{x^{2^N} - 1} \quad (x \neq 1).$$

²⁰⁴ For further reading, see:

- Cundy, H.M., *Mathematical Models*, Oxford University Press, 1961, 286 pp.
- Coxeter, H.S.M., *Regular Polytopes*, Dover, 1973, 321 pp.
- Fauvel, T. et al (eds), *Möbius and his band*, Oxford University Press, 1993, 172 pp.

no holes, and can be continuously deformed into a sphere. A convex polyhedron²⁰⁵ is said to be regular if its faces are regular and congruent polygons (e.g. cube, tetrahedron). The study of polyhedra held a central place in Greek geometry, which already recognized most of their salient geometrical features. Greek geometers correctly concluded that the only polygons that can occur as faces of a regular polyhedron are the regular polygons having 3, 4 or 5 sides, bringing the total number of possible regular polyhedra to five.

Now, all five of these possible forms actually exist. They were well known as early as **Plato** (ca 390 BCE), and he gave them a very important place in his *Theory of Ideas*, which is why they are often known as the “Platonic Solids”²⁰⁶. The most important data on the regular polyhedra are given in Table 4.3 (L = length of edge, R = radius of circumsphere).

While the sphere encloses the most volume of all shape having a given surface area, the tetrahedron, of all polyhedra, encloses the least volume with a given surface area [this ratio is equal to $(\frac{1}{12}a^3\sqrt{2})/a^2\sqrt{3} = \frac{a}{12}\sqrt{\frac{2}{3}}$, where a is the side length]. Table 4.3 suggests that for simple polyhedra $V - E + F = 2$, a fact first stated by **Descartes** (1635), proved incompletely by **Euler** (1751)

-
- Henle, M., *A Combinatorial Introduction to Topology*, Dover: New York, 1994, 310 pp.
 - Flegg, H.G., *From Geometry to Topology*, Dover: New York, 2001, 186 pp.

²⁰⁵ The designation ‘convex’ applies to every polyhedron that is entirely on one side of each of its faces, so that it can be set on a flat table top with any face down. Although convexity is *not* a topological property it *implies* a topological property, since every convex polyhedron is necessarily simple.

There is a peculiar difference between the convex and the non-convex polyhedra: whereas every closed convex polyhedron is rigid, there are closed non-convex polyhedra whose faces can be moved relative to each other.

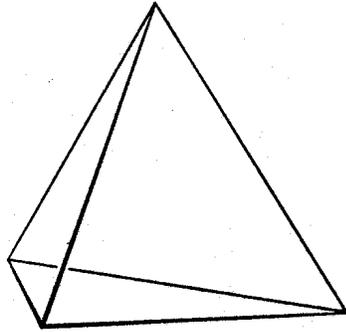
²⁰⁶ It seems probable that **Pythagoras** (c. 540 BCE) brought the knowledge of the cube, tetrahedron and octahedron from Egypt, but the icosahedron and the dodecahedron have been developed in his own school. He seems to have known that all five polyhedra can be inscribed in a sphere. These solids played an important part in Pythagorean cosmology, symbolizing the five elements: *fire* (tetrahedron), *air* (octahedron), *water* (icosahedron), *earth* (cube), *universe or earth* (dodecahedron). The Pythagoreans passed on the study of these solids to the school of Plato. Euclid discusses them in the 13th book of his *Elements*, where he proves that no other regular bodies are possible, and shows how to inscribe them in a sphere. The latter problem received the attention of the Arabian astronomer Abu al-Wafa (10th century CE), who solved it with a single opening of the compass.

for convex polyhedra, and proved generally by **Cauchy** (1811). It may have been known to **Archimedes** (ca 225 BCE), although the Greeks usually associated geometrical properties with measurements and not with mere counting.

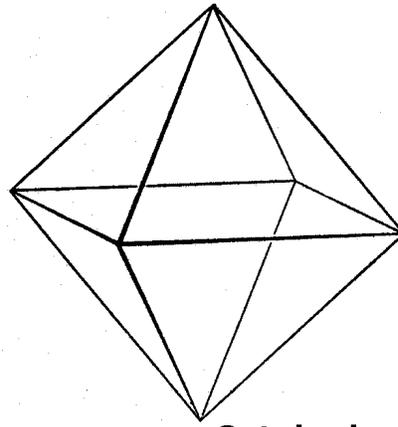
We have extant specimens of icosahedral dice that date from about the Ptolemaic period in Egypt. There are also a number of interesting ancient Celtic bronze models of the regular dodecahedron still extant in various museums. There was probably some mystic or religious significance attached to these forms. Since a stone dodecahedron found in northern Italy dates back to a prehistoric period, it is possible that the Celtic people received their idea from the region south of the Alps, and it is also possible that this form was already known in Italy when the Pythagoreans began to develop their teachings in Crotona.

Table 4.3 REGULAR POLYHEDRA

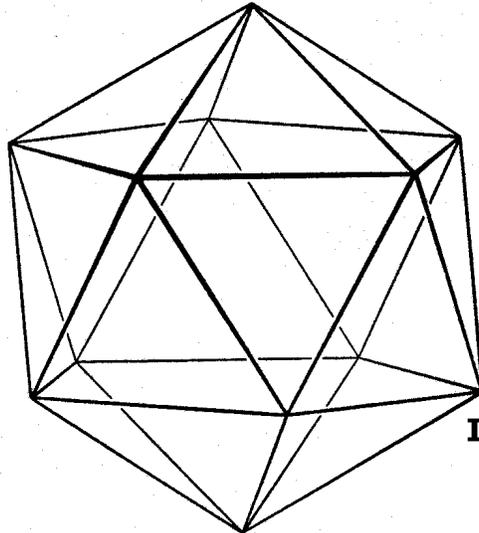
NAME OF POLYHEDRON	POLYGONS FORMING THE FACES	NUMBER OF				$\frac{R}{L}$
		VERTICES V	EDGES E	FACES F	FACES MEETING AT A VERTEX	
<i>Tetrahedron</i>	<i>Triangles</i>	4	6	4	3	$\frac{\sqrt{6}}{4}$
<i>Octahedron</i>	<i>Triangles</i>	6	12	8	4	$\frac{1}{\sqrt{2}}$
<i>Icosahedron</i>	<i>Triangles</i>	12	30	20	5	$\frac{1}{2}\sqrt{\frac{5+\sqrt{5}}{2}}$
<i>Cube (Hexahedron)</i>	<i>Squares</i>	8	12	6	3	$\frac{\sqrt{3}}{2}$
<i>Dodecahedron</i>	<i>Pentagons</i>	20	30	12	3	$\frac{1}{2}\sqrt{\frac{5+3\sqrt{5}}{2}}$



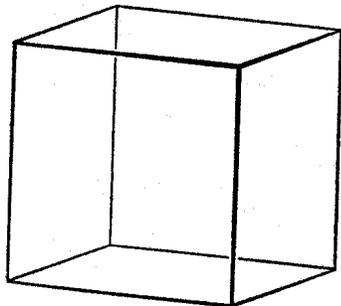
Tetrahedron



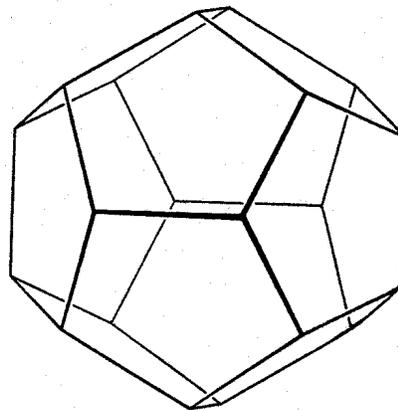
Octahedron



Icosahedron



Cube



Dodecahedron

The five regular polyhedrons attracted attention in the Middle Ages chiefly on the part of astrologers. At the close of this period, however, they were carefully studied by various mathematicians. Prominent among the latter was **Pietro Franceschi**, whose work *De Corporibus Regularibus* (c. 1475) was the first to treat the subject with any degree of thoroughness. Following the custom of the time **Pacoli** (1509) made free use of the works of his contemporaries, and as part of his literary plunder he took considerable material from this work and embodied it in his *De Divina Proportione*.

Albrecht Dürer, the Nürnberg artist, showed how to construct the figures from a net in the way commonly set forth in modern works.

Thus, Platonic and Archimedean polyhedra have sparked the imagination of creative individuals from Euclid to Kepler to Buckminster Fuller²⁰⁷. These polyhedra are rich in connections to the worlds of art, architecture, chemistry, biology, and mathematics. In the realm of life, the Platonic Solids present themselves in the form of microscopic organisms known as *radiolaria*.

Three other groups of polyhedra drew the attention of mathematicians throughout the ages:

- *Archimedean Solids*: characterized by having all their angles equal and all their faces regular polygons, not necessarily of the same species. Archimedes' own account of them is lost. Thirteen such solids exist mathematically, some realized in crystalline forms: truncated tetrahedron (8 faces); cuboctahedron (14); truncated cube (14); truncated octahedron (14); rhombicuboctahedron (26); icosidodecahedron (32); truncated icosahedron ($V = 60$, $E = 90$, $F = 32$); snub cube (38); rhombicosidodecahedron (62); snub dodecahedron (92). Recently, the truncated icosahedron showed up in chemistry as the molecule C_{60} , known as a *fullerene* (after **Buckminster Fuller**).
- *Kepler-Poinsot Polyhedra* have as faces congruent regular polygons, and the angles at the vertices all equal, but their center is multiply wrapped by the faces (convex polyhedra).

²⁰⁷ American engineer and inventor (1895–1983); among his numerous inventions is his *geodesic dome* structure (1947), based on 3-dimensional structural principles that were developed to achieve maximum span with a minimum material. His designs find parallels to such natural molecular geometries as the *tetrahedron* and the *truncated icosahedron* (C_{60} , named “Buckeyball” or “Fullerene” in his honor).

Fuller also built the geodetic dome at the American Pavillion in the 1970 World Fair in Montreal.

Four such solids exist: small stellated dodecahedron ($F = 12$, $V = 12$, $E = 30$); great dodecahedron ($F = 12$, $V = 20$, $E = 30$); great icosahedron ($F = 20$, $V = 12$, $E = 30$). They were described and studied by **Kepler** (1619), **Poinsot** (1810), **Cauchy** (1813) and **Cayley** (1859).

- *Semi-regular Polyhedra*: solids which have all their angles, faces, and edges equal, the faces not being regular polygons. Two such solids exist: rhombic dodecahedron, a common crystal form; and semi-regular triacontahedron.

On the basis of Euler's formula it is easy to show that there are no more than five regular polyhedra. For suppose that a regular polyhedron has F faces, each of which is an n -sided regular polygon, and that r edges meet at each vertex. Counting edges by faces, we see that

$$nF = 2E;$$

for each edge belongs to two faces, and hence is counted twice in the product nF ; but counting edges by vertices,

$$rV = 2E,$$

since each edge has two vertices. Hence from $V - E + F = 2$ we obtain the equation

$$\frac{2E}{n} + \frac{2E}{r} - E = 2$$

or

$$\frac{1}{n} + \frac{1}{r} = \frac{1}{2} + \frac{1}{E}.$$

We know to begin with that $n \geq 3$ and $r \geq 3$, since a polygon must have at least three sides, and at least three sides must meet at each polyhedral angle. But n and r cannot both be greater than three, for then the left hand side of the last equation could not exceed $\frac{1}{2}$, which is impossible for any positive value of E . Therefore, let us see what values r may have when $n = 3$, and what values n may have when $r = 3$. The totality of polyhedra given by these two cases yields the number of possible regular polyhedra.

For $n = 3$ the last equation becomes

$$\frac{1}{r} - \frac{1}{6} = \frac{1}{E};$$

r can thus equal 3, 4, or 5. (6, or any greater number, is obviously excluded, since $1/E$ is always positive.) For these values of r we get $E = 6, 12$, or 30 ,

corresponding respectively to the tetrahedron, octahedron, and icosahedron. Likewise, for $r = 3$ we obtain the equation

$$\frac{1}{n} - \frac{1}{6} = \frac{1}{E},$$

from which it follows that $n = 3, 4$, or 5 , and $E = 6, 12$, or 30 , respectively. These values correspond respectively to the tetrahedron, cube, and dodecahedron.

While Euler's formula is valid for simply-connected polyhedra (regular and truncated polyhedra, pyramids, prisms, cuboids, frustums, crystal-lattice unit cells of various kinds) which are all *topological spheres*, it fails for solids with holes in them and non-convex star-polyhedra. Thus, Kepler (1619) described the small and great *stellated dodecahedra* with $V = 12$, $F = 12$, $E = 30$, $V - E + F = -6$ and **Lhuilier** (1813) noticed that Euler's formula was wrong for certain families of solid bodies. For a solid with g holes Lhuilier showed that $V - E + F = 2 - 2g$.

Consider for example a non-simply-connected polyhedron such as the prismatic block, consisting of a regular parallelepiped with a hole having the form of a smaller parallelepiped with its sides parallel to the outer faces of the block. Introducing just enough extra edges and faces to render all faces simply-connected polygon interiors (rectangles and trapezoids), this polygon is seen to have $V = 16$, $E = 32$ and $F = 16$ such that $V - E + F = 0$. This corresponds to Lhuilier's formula with $g = 1$.

To understand the significance of the number g and its role in the topological classification of surfaces²⁰⁸, we compare the surface of the sphere with that of a torus. Clearly, these two surfaces differ in a fundamental way: on the sphere, as in the plane, every simple closed curve separates the surface into two disconnected parts. But on the torus there exist closed curves that do not separate the surface into two parts — for example, the two *generator circles* on the torus surface. Furthermore, such a closed curve cannot be continuously shrunk to a point — whereas *any* closed curve on a *sphere* can be so shrunk. This difference between the sphere and the torus marks the two surfaces as belonging to two topologically distinct classes, because this shows that it is impossible to deform one into the other in a continuous way.

Likewise, on a surface with two holes we can draw *four* closed curves each of which does not separate the surface into disjoint components; these can be

²⁰⁸ For the time being, we consider only *two-sided* and *closed* surfaces — i.e., we assume the surface has no boundary and that an ant, walking on one of its two sides, can never reach the opposite side without puncturing the surface. A 2-sided surface is also known as an *oriented* surface.

chosen to be the four generator curves (two per hole). Furthermore, one can draw *two* (non-intersecting) closed curves that, drawn simultaneously, still do not separate the two-hole surface. The torus is always separated into two parts by any two non-intersecting closed curves. On the other hand, *three* closed non-intersecting curves always separate the surface with two holes.

These facts suggest that we define the *genus* of a (closed and 2-sided) surface as the largest number of non-intersecting simple closed curves that can be simultaneously drawn on the surface without separating it. The genus of the sphere is 0, that of the torus is 1, while that of a 2-holed doughnut is 2. A similar surface with g holes has the genus g . The genus is a topological property of a surface and thus remains the same if the surface is deformed. Conversely, it may be shown that if two closed 2-sided (oriented) surfaces have the same genus, then one may be continuously deformed into the other, so that the genus $g = 0, 1, 2, \dots$ of such a surface characterizes it completely from the topological point of view.

For example, the two-holed doughnut and the sphere with two “handles” are both closed surfaces of genus 2, and it is clear that either of these surfaces may be continuously deformed into the other. Since the doughnut with g holes, or its equivalent, the sphere with g handles, is of genus g , we may take either of these surfaces as the topological representative of all closed oriented surfaces of genus g .

Suppose that a surface S of genus g is divided into a number of regions (faces) by marking a number of vertices on S and joining them by curved arcs. As stated above, it has been shown that

$$V - E + F = 2 - 2g,$$

where V = number of vertices, E = number of arcs, and F = number of faces or regions²⁰⁹. The topological invariant on the L.H.S. is usually denoted χ and is known as the *Euler characteristic* of the surface (this invariant admits a generalization to even-dimensional manifolds of dimension higher than two). We have already seen that for the sphere, $V - E + F = 2$, which agrees with the above equation, since $g = 0$ for the sphere.

Another measure of non-simplicity which is used in the classification of surfaces will emerge from the following example. Consider two plane domains: the first of these, a , consists of all points interior to a circle, while the second, b , consists of all points contained between two concentric circles. Any closed

²⁰⁹ An outline of the proof: S can be constructed from a particular partitioning of the sphere by *identifying* $2g$ distinct sphere faces pairwise. This reduces E and V by the same integer, and reduces F by $2g$, thus resulting in a reduction of $V - E + F$ from its sphere value (2), as claimed.

curve lying in the domain a can be continuously deformed or “shrunk” down to a single point *within the domain*. A domain with this property is said to be *simply connected*. The domain b , however, is not simply connected. For example, a circle concentric with the two boundary circles and midway between them cannot be shrunk to a single point within the domain, since during this process the curve would necessarily sweep through the center of the circles, which is not a point of the domain. A domain which is not simply connected is said to be *multiply connected*. If the multiply connected domain b is cut along a radius, the resulting domain is simply connected.

More generally, we can construct domains with two “holes”. In order to convert this domain into a simply connected domain, two cuts are necessary. If $h - 1$ non-intersecting cuts from boundary to boundary are needed to convert a given multiply connected planar domain D into a simply connected domain, the domain D is said to be h -tuply connected. The degree of connectivity of a domain in the plane is an important topological invariant of the domain. The number h is called the *connectivity number* assigned to every surface. It extends also, *mutatis mutandis*, to 3-dimensional bodies.

As an example, consider a closed, non-self-intersecting polygon (a *chain*) consisting of edges of a polyhedron. If the *surface* of the polyhedron is divided into two separate parts by every such closed chain of edges, we assign the connectivity $h = 1$ to the polyhedron. Clearly, all simple polyhedra have connectivity 1, since the surface of the sphere is divided into two parts by every closed curve lying on it. Conversely, it is readily seen that all polyhedra with connectivity 1 can be continuously deformed into a sphere. Hence the simple polyhedra are also called *simply connected*.

A polyhedron is said to have connectivity h if $h - 1$ is the greatest possible number of chains that, when simultaneously drawn, do not cut the surface in two. Since $h - 1 = 2$ for the prismatic block, its connectivity is $h = 3$.

We thus set $h = 1$ for the sphere and $h = 3$ for the torus. Surfaces of higher connectivity can be constructed by flattening a sphere made of a plastic material, cutting holes into it, and *identifying* (sewing together) each pair of stacked hole-boundary closed curves.

We shall call such surfaces *pretzels*. It can be proved that a pretzel with g holes (i.e. a g -handle surface) must have connectivity $h = 2g + 1$.

On a general surface, the curves can be chosen more freely than on a polyhedra, where we restricted the choice to chains of edges. Various other definitions can be given for the connectivity of surfaces – for example, the following:

On a closed surface of connectivity h , we can draw $h - 1$ closed curves without cutting the surface in two, but every system of h closed curves cuts the surface into at least two separate parts. On a closed surface of connectivity $h = 2g + 1$ there is at least one set of g closed, mutually non-intersecting curves – and no set of more than g such curves – having the property that the curves in the set do not cut the surface in two when drawn simultaneously.

All the polyhedra and closed surfaces we have considered thus far had odd connectivity numbers h and even Euler characteristics ($\chi = 2 - 2g$), related by the formula $\chi = 3 - h$. If we extend both concepts to surfaces with boundaries (i.e. open) — with χ still defined as $V - E + F$ and h now defined as the maximal number of simultaneous cuts (along closed or boundary-to-boundary open curves) leaving the surface connected — the formula becomes²¹⁰ $\chi = 2 - h$. And for such surfaces, χ and h may be both even or both odd.

The numbers χ , g and h are all topological invariants. So is the orientability/non-orientability property, which we explain next.

The question arises whether there are any closed (boundary-less) surfaces at all with even connectivities or odd χ values; or whether there are boundary-less surfaces for which genus and connectivity are not related by $h = 2g + 1$. Indeed, such surfaces do exist and are called one-sided (or non-orientable) surfaces.

Hitherto we have been dealing with “ordinary” surface, i.e. those having two sides. This restriction applied to closed surfaces like the sphere or the torus and to surfaces with boundary curves, such as the disk, a sphere with two holes (i.e. with two disks removed) – equivalent to a cylinder – or a torus from which a single disk has been removed.

The two sides of such a surface could be painted with different colors to distinguish them. If the surface is closed, the two colors never meet. If the surface has boundary curves, the two colors meet only along these curves. A bug crawling along such a surface and prevented from puncturing it or crossing boundary curves, if any exist, would always remain on the same side.

Möbius made the surprising discovery that there exist surfaces with only one side. The simplest such surface is the so-called Möbius strip (Figure 2), formed by taking a long rectangular strip of paper and pasting its two ends

²¹⁰ Also, the formula $h = 2g + 1$ does not always apply for a non-closed surface. For instance, a cylinder with g handles – equivalent to a g -handle sphere with two disks cut out – has $h = 2g + 2$; for $g = 0$ (a simple cylinder) $h = 2$, since it can be cut once ($1 = h - 1$) while maintaining connectedness — e.g. from boundary to boundary along the cylinder axis.

together after giving one end a half-twist. A bug crawling along this surface, keeping always to the middle of the strip, will return to its original position upside down and on the opposite side of the surface! The surface is thus indeed one-sided when considered globally; only local portions of it can be said to have two sides. The Möbius strip also has but one edge, for its boundary consists of a single closed curve. The ordinary two-sided surface formed by pasting together the two ends of a rectangle without twisting has two distinct, disconnected closed boundary curves; topologically it is a cylinder (or a sphere missing two disks).

If this surface is cut along a plane separating the two closed boundary-curves, it falls apart into two such disjoint cylinder surfaces, each with a new closed-curve component to its boundary. Like the cylinder, the Möbius strip has a continuous family of closed curves in its interior, each having the property of not being continuously deformable to a single point. And, as in the case of the cylinder, all such curves of unit winding-number (i.e. consisting of a single component if the surface is cut back into the original rectangle) can be deformed into each other, and are thus topologically equivalent.

However, unlike the cylinder, if the Möbius strip is cut along one of its non-shrinkable closed curves, we find that it remains in one piece²¹¹. It is rare for anyone not familiar with the Möbius strip to predict this behavior, so contrary to one's intuition of what "should" occur. If the surface that results from cutting the Möbius strip along the middle is again cut along its middle, two separate but intertwined strips are formed.

The connectivity of the Möbius strip is $h = 2$, just as the untwisted open cylinder. It also may be characterized by means of another important topological concept which can be formulated as follows: Imagine every point of a given surface (with the exception of boundary points, if any) to be enclosed in a small closed curve that lies entirely on the surface. We then try to fix a certain sense (handedness) on each of these closed curves in such a way that any two curves that are sufficiently close together have the same sense. If such a consistent determination of sense of traversal is possible in this way, we call it an *orientation* of the surface and call the surface *orientable*.

While all two-sided surfaces are orientable, one-sided surfaces are not. Thus the classification of surfaces into two-sided and one-sided surfaces is identical to the classification into orientable and non-orientable surfaces.

²¹¹ The cut strip is in fact equivalent to a rectangular strip subjected to *two* half-twists before identifying its two (short) opposite sides — both half-twists being in the same sense. This strip is topologically equivalent to a cylinder, yet cannot be deformed into it without self-intersection if embedded in 3-D space (\mathbb{R}^3).

It is easy to see that a surface is non-orientable if and only if there exists on the surface some closed curve such that a continuous family of small oriented circles whose center traverses the curve will arrive at its starting point with its orientation reversed.

The Möbius strip is an open one-sided surface and does not intersect itself. But it can be proven that all one-sided closed surfaces embedded in \mathbb{R}^3 (Euclidean 3-dimensional space) have self-intersections. However, the presence of curves of self-intersection need not represent a topological property in the sense that in some cases it can be transformed away by deformation, or eliminated by defining the surface *intrinsically* (without embedding it in a 3-D \mathbb{R}^3 space), or else by embedding it in an \mathbb{R}^n space with $n > 3$. If this is not the case we say that the surface has *singular points* which are a topological property.

This raises the question of whether there can exist any one-sided closed surface (2-D intrinsic manifold) that has no singular points. Such a surface was first constructed mathematically by Felix Klein, as follows. Consider an open tube (cylinder). A *torus*²¹² is obtained from it by bending the tube until the ends meet and then gluing (identifying) the boundary circles together. But the ends can be welded in a different way:

Taking a tube with one end a little thinner than the other, we bend the thin end over and push it through the wall of the tube, molding it into a position where the two circles at the ends of the tube have nearby and concentric positions. We now expand the smaller circle and contract the larger one a little until they meet, and then join them together (Fig. 7). This does not create any singular points and gives us Klein's surface, also known as the *Klein bottle*. It is clear that the surface is one-sided and, in any \mathbb{R}^3 embedding, intersects itself along a closed curve where the narrow end was pushed through the wall of the tube.

The connectivity number of the Klein bottle is 3, like that of a torus. It can be shown that any closed, one-sided surface of genus g is topologically

²¹² *Torus*: a surface (intrinsic or embedded in \mathbb{R}^3). The *intrinsic* torus is a rectangle with opposite ends identified without twists (Fig. 9(f)). An \mathbb{R}^3 -embedded torus is generated by revolving a circle about a line (in its plane) that does not intersect the circle. One of its parametric representations in Gaussian surface coordinates (u, v) is

$$\mathbf{r}(u, v) = [(a + b \cos v) \cos u; (a + b \cos v) \sin u; b \sin v], \\ a > b > 0; \quad 0 \leq u < 2\pi, \quad 0 \leq v < 2\pi.$$

a and b are the two radii of the \mathbb{R}^3 torus, while the coordinates u, v are azimuths along two generating circles.

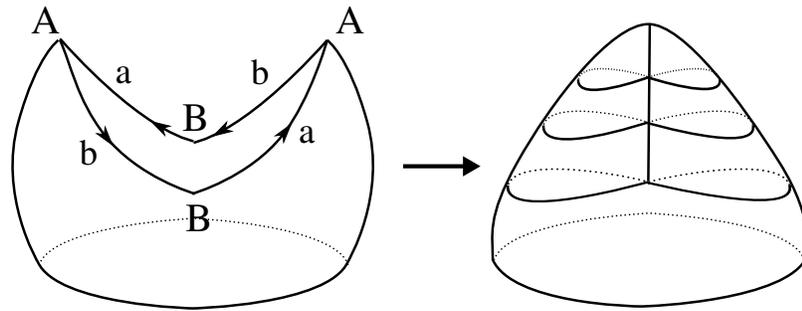


Fig. 1: Sewing up a cylinder to yield a representation of the Möbius strip as a topological sphere with cross-cap. The two copies of point A are identified, and similarly for B and the directed arcs a, b .

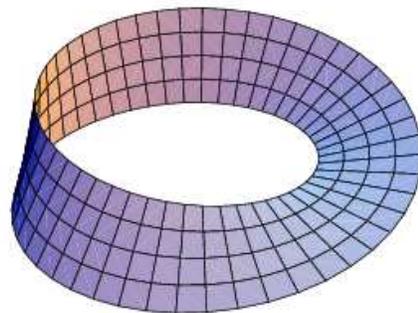


Fig. 2: An embedding of the Möbius strip in \mathbb{R}^3 .

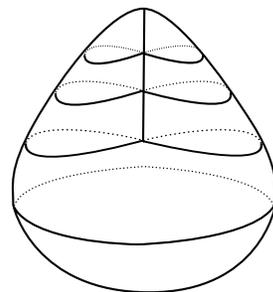


Fig. 3: The Real Projective Plane (sphere with one cross-cap).

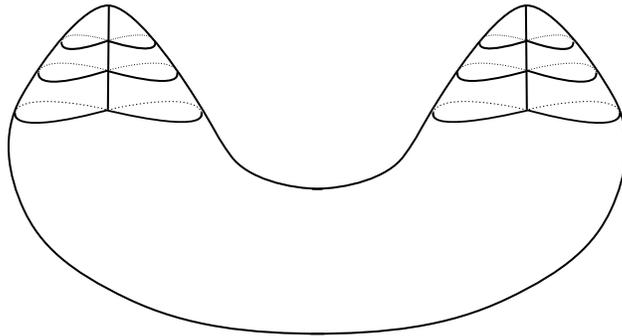


Fig. 4: Klein bottle represented as a topological sphere with two cross-caps.

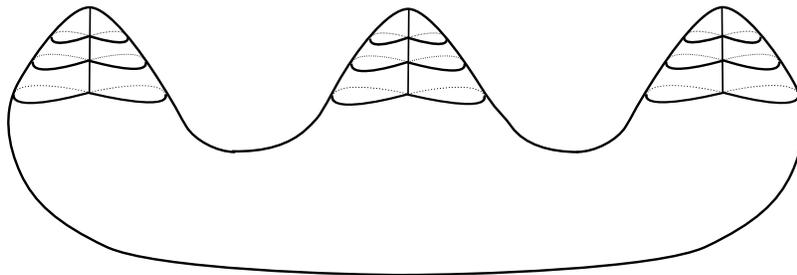


Fig. 5: The triple-crosscap surface (topological sphere with three cross-caps).

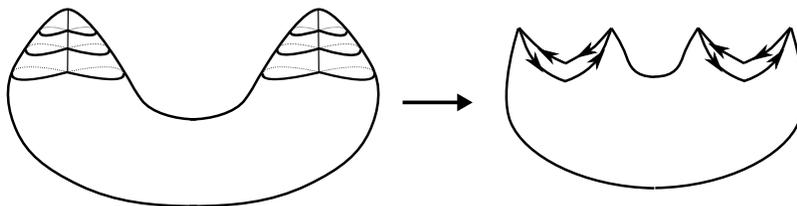


Fig. 6: Cutting the Klein bottle along to closed curves while maintaining connectivity.

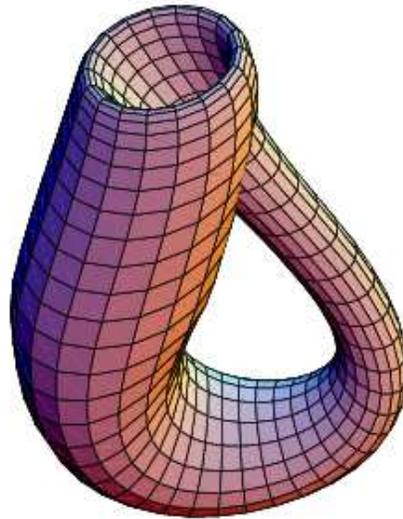
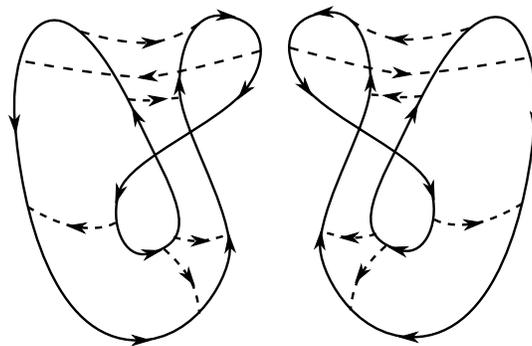
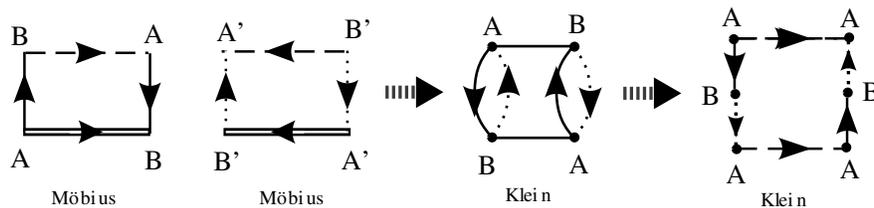


Fig. 7: A Klein bottle represented as a self-intersecting, nonsingular embedding (*immersion*) in \mathbb{R}^3 .



(a) Two Möbius strips with their boundaries sewn (identified) yield a Klein bottle.



(b)

Fig. 8

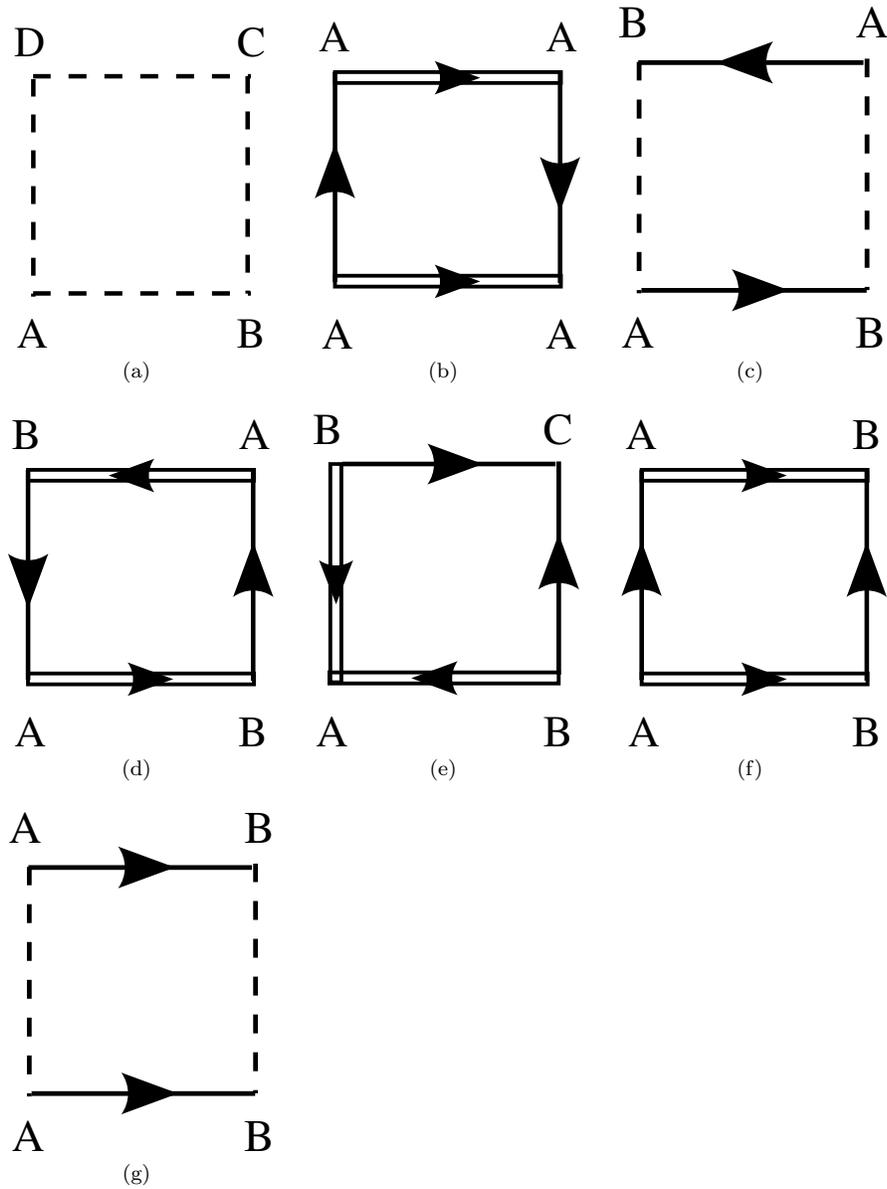


Fig. 9: Surfaces (open and closed, orientable and non-orientable) obtained from a rectangle by identifying (sewing) edges in various ways. Broken-line edges are not identified; arrows (single-line or double lines) are identified with same-type arrows (head with head and tail with tail). (a): topological disk; (b): Klein bottle; (c): Möbius strip; (d): real projective plane; (e): sphere; (f): torus; (g): cylinder (tube).

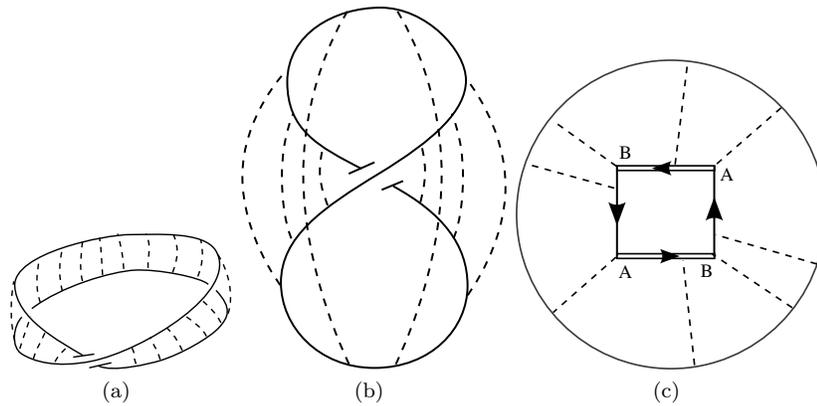


Fig. 10: Deforming the *standard* Möbius-strip representation into a *disk-with-crosscap*.

equivalent to a sphere from which g disks have been removed and replaced by local topological constructs called *cross-caps*.²¹³

²¹³ Cross-Caps: Modular building blocks for non-oriented surfaces.

The Möbius strip is the basic unit of non-oriented surfaces. It can be represented as described above: by half-twisting a rectangle and then identifying a pair of opposite sides (Figs. 2 and 9(c)). In the latter, the two corners labeled “A” are identified – “sewn” together – as are the two corners labeled “B”; and the two solid edges are identified in accordance with the arrow-indicated directions. But in this representation the Möbius strip has a complicated-looking boundary. Figures 1 and 10(a) to 10(c) show how to continuously deform this into the standard disk-with-crosscap representation. Unfolding the solid-line closed boundary of 10(a) yields 10(b) (the dotted lines indicate the Möbius-strip interior). Then, untwisting the boundary of 10(b) into a circle yields Fig. 10(c). In 10(c) it was necessary to cut the surface along some closed curve ABA to avoid intersections among the broken lines. This cut results in the rectangle depicted in 10(c), in which the two single-solid lines are identified along their arrows — as are the two double-solid lines. Re-sewing the cut ABA , as depicted in Fig. 1, results in a disk with one crosscap. This is a convenient representation of the Möbius strip, because its boundary is a simple curve (a circle if we wish), and also because any non-orientable surface (open or closed) is topologically equivalent to a sphere with some number of local disks removed, with some or all of these disks replaced by cross-caps. On the other hand — as clearly seen in Fig. 1 — the cross-cap representation of the Möbius strip makes it self-intersecting in a 3-D embedding (though it has no singular point).

From this it easily follows that the Euler characteristic $V - E + F$ of such a surface is related to g by the equation

$$V - E + F = 2 - g.$$

Thus for the Klein bottle, $g = 2$. For an orientable closed surface we have $\chi = 2 - 2g$, while for a non-orientable closed surface $\chi = 2 - g$.

Figs. 3-5 depict the three simplest classes of *closed* non-orientable surfaces, represented as a sphere with one, two and three local cross-caps, respectively. The class shown in Fig. 3 includes the *real projective plane* ($\mathbb{R}P^2$), obtained from \mathbb{R}^3 by identifying all points $(\lambda x, \lambda y, \lambda z)$ with fixed (x, y, z) and all real numbers λ ; this class is also a Möbius strip with its boundary sewn to a disk. Fig. 4 — a sphere with two cross-caps — is equivalent to a Klein bottle (Fig. 7). This is because a Klein bottle can be constructed by sewing together the boundaries of two Möbius strips — as shown in two different ways in Figs. 8(a) and 8(b). Fig. 8(a) shows how two standard (twisted-strip) representations of Möbius strips are sewn along their boundaries to yield a single Klein bottle. In Fig. 8(b) it is done another way, by identifying the boundaries of two Möbius strips. Each separate Möbius strip is represented as a rectangle with two of its edges identified, as in Fig. 9(c). The final Klein bottle can be represented as a rectangle with its four edges identified pairwise as in Fig. 9(b). The vertices A, A' are identified with each other, as are B and B' ; the two single-broken-line arrows are identified with each other (base with base and arrow-tip with arrow-tip), and the two double-solid-line arrows are similarly identified. The two remaining arrow pairs are separately identified *within* each Möbius strip, as in Fig. 9(c) (single-solid-arrows identified with each other, as are the single-dotted-arrows). Proceeding from left to right, the first solid-white arrow indicates the sewing together of the boundaries of the two Möbius strips. The second solid-white arrow indicates two further operations: a 180° twisting of the *right* closed curve ABA to align it with the *left* closed ABA curve, followed by a cut along a curve between the two copies of point A . Neither operation changes the Klein bottle's topology.

Fig. 5 shows a sphere with three cross-caps; this can be shown to be topologically equivalent to a torus with a small disk removed and replaced with a cross-cap (*Dyck's theorem*).

Fig. 6 shows how a Klein bottle can be simultaneously cut along two closed curves while remaining a connected surface: the two cuts open the surface's two cross-caps, converting them into two closed-curve boundaries — the Klein bottle is thus converted into a topological cylinder.

Finally, Figs. 9(a)-9(g) show how to obtain various surfaces — open and closed, orientable and one-sided — by sewing (identifying) the vertices and edges of a single rectangle in various ways.

In general, for a closed orientable surface S of genus g , for which the $\mathbb{R}^2 \rightarrow \mathbb{R}^3$ mapping $\mathbf{r}(u, v)$ is twice continuously differentiable, the integral curvature is equal to $\iint_S K dA = 4\pi(1 - g)$, where K is the Gaussian curvature (this follows from the Gauss-Bonnet theorem). For the torus ($g = 1$) the right-hand side vanishes. And indeed, in terms of the parametric representation of the \mathbb{R}^3 -embedded torus given above:

$$K = \frac{\cos v}{b(a + b \cos v)}, \quad dA = (a + b \cos v)b du dv$$

and therefore the integral curvature is zero as claimed, as $\int_0^{2\pi} dv \cos v = 0$. Another interesting feature of the torus is that an elliptic function defines a mapping of a plane into a torus. It arises from the notion that the curve $y^2 = ax^3 + bx^2 + cx + d$ can be parametrized as $x = f(z)$, $y = f'(z)$, where f and f' are elliptic functions (**Jacobi**, 1834).

II. TOPOLOGICAL MAPPINGS

Starting with the concept of a *set* (such as points in a plane, lines through a point, rotations in 3-D, etc.) one can generalize the idea of a *function* to that of a *map*: A map is a relation between two specified sets that associates a unique element of the second to each element of the first.

To establish *topological equivalence* between sets, one must have a mathematical machinery that is able to transform one of the sets into the other, and this transformation must be a map endowed with various properties.

There are various methods of mapping one surface (or higher-dimensional manifold space, whether intrinsically defined or embedded) onto another. The most faithful image of a surface is obtained by an *isometric*, or *length-preserving*, mapping²¹⁴. Here the geodesic distance between any two points is preserved (assuming the surface or space is endowed with a *metric*²¹⁵), all angles remain unchanged, and geodesic lines are mapped into geodesic lines. An isometry also preserves the Gaussian curvature at corresponding points. Hence the only surfaces that can be mapped isometrically into a part of the plane are surfaces whose Gaussian curvature is everywhere zero; this excludes, for example, any portion of the sphere. In consequence, no *geographical map* (i.e. map of the earth's surface) can be free of distortions.

Less accurate, but also less restrictive, are the *area-preserving mappings*. They are defined by the condition that the area enclosed by every closed curve be preserved. With the aid of such a mapping portions of the sphere can be mapped onto portions of the plane, and this is frequently used in geography. It is achieved in practice by projecting points of the sphere onto the cylinder along the normals of the cylinder. If the cylinder is now slit open along a generator and developed into a plane, the result is an area-preserving image of the sphere in a plane; the distortion increases the further we are from the circle along which the cylinder touches the sphere.

Another type of mapping, especially useful for navigation, is that of *geodesic maps*, where geodesics are preserved. If, for example, a portion of

²¹⁴ In the Euclidean plane all isometries can be generated by combining a two-parameter translation, a one-parameter rotation, and a single reflection about some fixed axis. In Euclidean \mathbb{R}^3 , there are 3 translations parameters and 3 rotation-angle parameters, but still only one independent reflection, which can be implemented with the help of a *mirror*. No more than 3 mirrors (i.e. three reflection planes) are needed to generate any isometry in \mathbb{R}^3 .

²¹⁵ In the case of a 2D surface embedded in \mathbb{R}^3 , the natural surface metric is the one inherited from the Euclidean metric of the "host" \mathbb{R}^3 space.

a sphere is projected from its center onto a plane, then the great circles are mapped into straight lines of the plane, and the map is therefore geodesic. At the same time, this gives us a (local) geodesic mapping of all surfaces of constant positive Gaussian curvature into the plane, because all these surfaces can be mapped isometrically into spheres. All surfaces with constant negative Gaussian curvature can also be mapped into the plane by a geodesic mapping.

Yet another type of mapping is that of the *conformal*, or *angle-preserving*, mappings, for which the angle at which two curves intersect is preserved. The simplest examples of conformal mapping, apart from the isometric mapping, are *stereographic projections* and the *circle-preserving transformations*²¹⁶. A stereographic projection map, in which a sphere (with its north-pole removed) is placed atop a plane and projected onto it by drawing straight lines from that pole, is also a circle-preserving map.

It can be shown that *very small figures* suffer hardly any distortion at all under general conformal transformations; not only angles are preserved, but the *ratios of distances* (although not the distances themselves) are approximately preserved. *In the small*, the conformal mappings are thus the nearest thing to isometric mappings among all the types of mappings mentioned earlier, for area-preserving and geodesic mappings may bring about arbitrarily great distortions in arbitrarily small figures.

The most general mappings that are at all comprehensible to visual intuition are *continuous invertible mappings (homeomorphisms)*. The only condition here is that the mapping is *one-to-one* and that neighboring points (and *only* such) go over to neighboring points. Thus a homeomorphic mapping may subject any figure to an arbitrary amount of distortion, but it is not permitted to *tear* connected regions apart or to *stick* separate regions together. Yet, continuous mappings do not always exist that can map (which we refer to here as “continuous” for simplicity’s sake) two given surfaces onto each other (Example: the circular disk and the plane annulus bounded by two concentric circles cannot be mapped continuously into each other, even not their boundaries alone!). Clearly, the class of continuous mappings embraces all the types of mapping mentioned so far. The question of when two surfaces can be mapped onto each other by a continuous mapping is one of the problems of topology.

The simplest type of topological mapping of a surface consists of a continuous mapping (homeomorphism) which is such as to transform the surface

²¹⁶ An example of a circle-preserving map is the $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ *inversion map* w.r.t. a given circle. If the latter is $x^2 + y^2 = a^2$, this inversion is $x \rightarrow \frac{a^2 x}{x^2 + y^2}$, $y \rightarrow \frac{a^2 y}{x^2 + y^2}$. It is a special type of conformal transformation that maps *any* circle into another circle.

as a whole onto itself, and which is arcwise-connected with the trivial (unity) map (in which each surface point is mapped to itself²¹⁷). This type of mapping is called a *deformation*. The reflection of a plane about a straight line, on the other hand, is an example of a topological mapping which is not a deformation; for a reflection reverses the sense of traversal (orientation) of every circle, whereas deformations cannot reverse the sense of traversal.

A point that is mapped onto itself under the mapping is called a *fixed point* of the mapping. In the applications of topology to other branches of mathematics, “fixed-point” theorems play an important role. The *theorem of Brouwer* states that every continuous deformation of a circular disk (with the points of the circumference included) onto itself has at least one fixed point. On a sphere, any continuous transformation which carries no point into its diametrically opposite points (e.g. any *small deformation*) has a fixed point.

Fixed point theorems provide a powerful method for the proof of many mathematical “existence theorems” which at first sight may not seem to be of a geometrical character. Also, topological methods have been applied with great success to the study of the qualitative behavior of dynamical systems. A famous example is a fixed point theorem conjectured by **Poincaré** (1912), which has an immediate consequence: the existence of an infinite number of *periodic orbits* in the restricted problem of three bodies.

Apart from the choice of the mapping transformation there is yet another problem that must be resolved; in describing a surface or other manifold, there is the freedom of choice of a suitable *coordinate system* (CS). In general we cannot restrict ourselves to manifolds which can be covered by a *single CS* such as is suitable for an n -dimensional Euclidean space \mathbb{R}^n ; simple examples of various kinds of surfaces embedded in E^3 indicate that, in general, no single CS can exist which covers a given surface completely.²¹⁸

The simplest example is a 2-dimensional spherical surface in E^3 (the latter having Cartesian coordinates (x_1, x_2, x_3)), which we wish to map onto a planar

²¹⁷ Two continuous mappings $f : A \rightarrow B, g : A \rightarrow B$ from a set A to a set B are said to be arcwise-connected (or *continuously deformable* into each other) if there exists an arc of continuous functions $h(s) : A \rightarrow B, s \in [0, 1]$, such that $h(0) = f, h(1) = g$, and h is a continuous map from $[0, 1] \times A$ to B .

²¹⁸ E^n is the space \mathbb{R}^n with a Euclidean metric (norm). A CS (coordinate system) is a homeomorphism between a region (open subset) of the surface (or higher-dimensional manifold) and \mathbb{R}^m , m being the manifold’s *dimension* ($m = 2$ for a surface). For a manifold requiring more than one CS, it is assumed that the open subsets *cover* the manifold, and that in the intersection of any two subsets, the two CS maps compose to yield a $\mathbb{R}^m \rightarrow \mathbb{R}^m$ homeomorphism.

disc. To obtain a one-to-one correspondence in the mapping, one may choose the hemisphere for which $x_1 > 0$, which is then continuously mapped onto a disk in the x_2x_3 plane. Accordingly, this hemisphere is referred to as a *coordinate neighborhood*.

Similarly, 5 other hemispheres corresponding to the respective restrictions $x_1 < 0$; $x_2 > 0$; $x_2 < 0$; $x_3 > 0$; $x_3 < 0$ can be regarded as coordinate neighborhoods. The totality of these 6 hemispheres covers the sphere completely, and in the overlap of any pair of them, composing the two corresponding maps yields a continuous map of one planar disk onto another. In general, the existence and overlap structure of suitable coordinate neighborhoods depends on the topological properties of the surface taken as a whole. This shows that one must give up on the construction of a unique CS for all points of a space under consideration and use different CS for different parts of the space.

A surface, however curved and complicated, can be thought of as a set of little curved patches glued together; and topologically (though not geometrically) each patch is just like a patch in the ordinary Euclidean plane. It is not this *local* patch-like structure that produces things like the hole in a torus: it is the *global* way all the patches are glued together. Once this is clear, the step to n dimensions is easy: one just assembles a space from little patches carved out of n -dimensional space instead of a plane. The resulting space is an *n -dimensional manifold*. For example: the motion of three bodies under mutual gravitational forces involves an 18-dimensional phase-space manifold, with 3 position coordinates and 3 velocity coordinates per body.

III. ALGEBRAIC TOPOLOGY

Algebraic topology is the study of the *global* properties of spaces by means of abstract algebra. One of the earliest examples is *Gauss's linkage formula* which tells us whether two closed space curves are linked, and – if so – how many times does any one of them wind around the other. The linkage number remains the same even if we continuously deform the space curves. The central idea here is that continuous geometric phenomena can be understood by the use of integer-valued topological invariants.

One of the strengths of algebraic topology has always been its wide degree of applicability to other fields. Nowadays that includes fields like theoretical physics, differential geometry, algebraic geometry, and number theory. As an example of this applicability, here is a simple topological proof that every

non-constant polynomial $p(z)$ has a complex zero (root) — a key component in proving the fundamental theorem of algebra.

Consider a circle of radius R and center at the origin of the complex plane. The polynomial transforms this into another closed curve in the complex plane. If this image curve ever passes through the origin, we have our zero. If not, suppose the radius R is very large. Then the highest power of $p(z)$ dominates and hence $p(z)$ transforms the circle into a curve which winds around the origin the same number of times as the degree of $p(z)$. This is called the *winding number* of the curve around the origin. It is always an integer and it is defined for every closed curve which does not pass through the origin. If we deform the curve, the winding number has to vary continuously but, since it is constrained to be an integer, it cannot change and must be a constant unless the curve is deformed through the origin.

Now deform the image curve by shrinking the radius R to zero and suppose that the image curve never passes through the origin, that is to say, the original circle, in shrinking, never passes through a zero of the polynomial. The image curve gets very small since $p(z)$ is continuous; hence it must have winding number 0 around the origin unless it is shrinking to the origin (which cannot be the case unless $p(0) = 0$). If the image curve is shrinking to the origin, the origin is a zero of $p(z)$. If not, the winding number is 0 which means that the polynomial must have degree 0; in other words, it is a constant.

The winding number of a curve illustrates two important principles of algebraic topology. First, it assigns to a geometric object, the closed curve, a discrete invariant, the winding number which is an integer. Second, when we deform the geometric object, the winding number does not change, hence, it is called an invariant of deformation or, synonymously, *an invariant of homotopy*. The field is called *algebraic topology* because an equivalence class of geometric entities possessing the same invariant — e.g. linkage number between curves; winding numbers of curves about points, or of closed surfaces in many-to-one mappings about other closed surfaces; winding numbers of non-shrinkable generator curves on the surface of a surface of nonvanishing genus; et cetera — turn out to form algebraic structures, such as rings and groups, under various geometric operations.

IV. FROM CURVES AND KNOTS TO MANIFOLDS

A simple closed curve (one that does not intersect itself) is drawn in the plane. What property of this figure persists even if the plane is regarded as

a sheet of rubber that can be deformed in any way? The length of the curve and the area that it encloses can be changed by a deformation. But there is a topological property of the configuration which is so simple that it may seem trivial: A simple closed curve C in the plane divides the plane into exactly two domains, an inside and an outside. By this is meant that those points of the plane not on C itself fall into two classes — A , the outside of the curve, and B , the inside — such that any pair of points of the same class can be joined by a curve which does not cross C , while any curve joining a pair of points belonging to different classes must cross C . This statement is obviously true for a circle or an ellipse, but the self-evidence fades a little if one contemplates a complicated curve like the twisted polygon. This problem was first stated by **Camille Jordan** (1882) in his *Cours d'analyse*. It turned out that the proof given by Jordan was invalid.

The first rigorous proofs of the theorem were quite complicated and hard to understand, even for many well-trained mathematicians. Only recently have comparatively simple proofs been found²¹⁹. One reason for the difficulty lies in the generality of the concept of “simple closed curve”, which is not restricted to the class of polygons or “smooth” curves, but includes all curves which are topological images of a circle. On the other hand, many concepts such as “inside”, “outside”, etc., which are so clear to the intuition, must be made precise before a rigorous proof is possible.

It is of the highest theoretical importance to analyze such concepts in their fullest generality, and much of modern topology is devoted to this task. But one should never forget that in the great majority of cases that arise from the study of concrete geometrical phenomena it is quite beside the point to work with concepts whose extreme generality creates unnecessary difficulties. As a matter of fact, the Jordan curve theorem is quite simple to prove for the reasonably well-behaved curves, such as polygons or curves with continuously turning tangents, which occur in most important problems.

A *knot* is formed by first looping and interlacing a piece of string and then joining the ends together. The resulting closed curve represents a geometrical figure the “knotiness” of which remains essentially the same even if it is deformed by pulling or twisting without breaking the string. But how is it possible to give an intrinsic characterization that will distinguish a knotted closed curve in space from an unknotted curve such as the circle? The answer is by no means simple, and still less so is the complete mathematical analysis of the various kinds of knots and the differences between them. Even for the simplest case this has proved to be a daunting task.

²¹⁹ A generalization of the *Jordan theorem* to arbitrary surface is used in proving the *surface classification theorem* cited earlier.

Consider, for example, two knots which are completely symmetric mirror images of one another. The problem arises whether it is possible to deform one of these knots into the other in a continuous way. The answer is in the negative, but the proof of this fact requires considerable knowledge of the technique of topology and group theory.

Knots are the most immediate topological features of curves in space. Beyond curves come surfaces; beyond surfaces come multidimensional generalizations called *manifolds*, introduced by **Riemann**.

Whereas mathematical analysis and the theory of differential equations deal primarily with “local” properties of a function (only infinitesimally adjacent points are considered), geometry studies the “global” properties of functions (i.e. their properties are analyzed by considering finitely spaced points). This intuitive idea of globality has given rise to the fundamental concept of *manifold* as a generalization of the concept of *domain* in Euclidean space.

A coordinate system describing the positions of points in space is an indispensable tool for studying geometrical objects. Using coordinate systems, we can apply the methods of differential and integral calculus to solve various problems. Therefore, an analysis of spaces which admit such concepts as differentiable or smooth functions, differentiation and integration, has emerged as an independent branch of geometry.

Topologists would like to do for manifolds what they have already done for surfaces and knots. Namely:

- (1) Decide when two manifolds are or are not topologically equivalent.
- (2) Classify all possible manifolds.
- (3) Find all the different ways to embed one manifold in another (e.g. a knotted circle in 3-space).
- (4) Decide when two such embeddings are, or are not, the same.

The answer to problems (1) and (2) lies in an area called *homotopy theory* which is part of *algebraic topology*. It endeavors to associate various algebraic invariants with topological spaces. **Poincaré** was one of the fathers of this theory. But problems (1) and (2) have not yet been fully resolved. Problem (3) led topologists to some surprising and counter-intuitive results, as the following example shows. It has been asked: when can two 50-dimensional spheres be *linked* [i.e. embedded such that they cannot be separated by a topology-preserving transformation of the surrounding n -dimensional space]. The answer is:

cannot link for $n \geq 102$
 can link for $n = 101, 100, 99, 98$
 cannot link for $n = 97, 96$
 can link for $n = 95, \dots, 52$.

V. NETWORKS

Graph (or Network) theory had its origin in a paper by Euler (1736) including the famous problem of the *bridges of Königsberg*. **Euler** saw that the problem could more easily be studied reducing island and banks to *points* and drawing a *network* (graph) in which two points are connected by an *edge* whenever there is a bridge connecting the corresponding two land masses. In this way Euler was able to abstract the problem so that only information essential to solving it was highlighted, and he could dispense with all other aspects of the problem. He could thus rephrase the problem as follows: “Given a connected graph, find a path that traverses each edge of the graph without retracing any edge.” Such a path is called a *eulerian traversal* or *eulerian path*²²⁰. Some experimentation and application of logic lead to the conclusion that in order to have a eulerian path, it is necessary that for any edge along which the path enters a vertex, there must correspond a distinct edge along which the path leaves it — and that all such edge-pairs be distinct for any given vertex. The only exception occurs for the beginning and ending vertices of the path, if these points are different.

Networks can be used to solve *mazes* and guarantee that one can find a path through a maze, if such a path exists, even when no map is explicitly given. Other procedures enable people to retrace their steps to the beginning of a *labyrinth*. Some of these procedures have applications to problems of computer processing, traffic control, electrical engineering, and many other fields.

During the ten generations elapsed since 1736, mathematicians have developed a new branch of geometry — a *geometry of dots and lines*, otherwise known as *graph theory* — that preserves geometrical relations only in their most general outlines. Here lines do not have to be straight, nor are there such things as perpendicular or parallel lines, and it does not make sense to talk about bisecting lines or measuring lengths or angles. The power of graph theory (a sub-field of topology) is that it can be used to model many patterns

²²⁰ A *practical architectural application*: In the hallways of a museum, pictures are hung on one side of each hall. How does one design a tour that will enable a person to see each exhibit exactly once?

in nature — from the branching of rivers to the cracking or brittle of surfaces to subdivision of cellular forms, as well as many abstract concepts. It gives us a way to study spatial structures unencumbered by the details of Euclidean geometry.

A *geographical map* shows countries, borders and corners. From such a map we may prepare an abstract *mathematical map* in which countries are *faces* (F), borders are chains of pairs of adjacent *edges* (E) and corners are *vertices* (V). In order to study the topology of a map in the technical language of mathematics, we must forget its geographical significance and treat it as merely a network, or graph, being a set of faces, edges and vertices, $M = \{F, E, V\}$, with certain *incidence relations* among them (e.g. face f_1 has edges (e_2, e_3) ; edge e_2 is shared by faces $\{f_1, f_3\}$; vertex v_1 is shared by $\{f_1, f_2, f_3\}$ and also by $\{e_2, e_3, e_4\}$; etc.). In this context the face is represented by some polygon and each edge lies in exactly two faces. Copies of a map formed by placing it on a flexible membrane and stretching the membrane without cutting, are considered identical or *homeomorphic*. Edges and faces thus become distorted but the sets E , F and V and their relational structure (incidence relations) maintain their integrity.

From a mathematical point of view, maps on a plain and maps on the sphere, with one point removed, are isomorphic. Since all the enclosed areas, including one additional outer one, are now considered to be faces, and maps are always considered to be in one piece (connected), one can show that *Euler's formula* $V - E + F = 2$ holds for connected planar maps on either a plane or a sphere.

There is a family of maps for which each vertex, edge or face is like every other vertex, edge or face. They are called *regular maps* and are said to have *perfect symmetry*. Upon finding oneself stranded in a mathematical country defined by such a map, one would experience vistas of sameness in all directions and be hopelessly lost.

There are only five²²¹ such regular maps and they correspond to the five 3-dimensional Platonic Solids. In fact, they are obtained by projecting the edges of a Platonic polyhedron onto a plane from a point directly above the center of one of its faces, and counting the infinite area outside the boundary as an additional face. These are known as *Schlegel diagrams*.

Visually, this amounts to holding one face of a polyhedron quite close to one's eyes, looking at the structure through the face, and drawing the projection of the structure as seen in this exaggerated perspective. The number of vertices, faces and edges for the Schlegel diagrams then becomes identical to those of the corresponding Platonic Solids.

²²¹ Except for two *trivial* families, one of which consist of all regular polygons.

Just as there are only *five* regular maps on the sphere (or plane), there are only *three* classes of regular maps that can be created on a torus.

For a surface homeomorphic to a sphere with g handles Euler's formula becomes $V - E + F = 2 - 2g$.

On a torus, for example, we have (with $g = 1$) $V - E + F = 0$.

G. R. Kirchhoff enunciated (1845) laws which allow calculations of currents, voltages and resistances of *electrical networks*. In the framework of these laws he became interested in the mathematical problem of the number of independent circuit equation in a given network. Considering the electrical network as a geometrical object (map) constructed from points (vertices, V) and lines (edges, E), Kirchhoff proved that, in general, the number of independent circuits²²² is equal to $(E - V + 1)$.

His paper is quite modern in its approach, and he used various constructions which we now think of as standard in graph theory. But he did not have the algebraic techniques that are needed to extend the results to higher dimensions. However, the basic ideas were latent in Kirchhoff's paper, and it was just those ideas which mathematicians were able to develop in the second half of the 19th century, in order to create what we now call '*algebraic topology*'. This development did not happen overnight.

The apparatus of vectors, matrices, and what we call now *linear algebra*, as well as the *abstract algebra* of groups, rings, homeomorphisms etc., were not available to Kirchhoff, Listing, and the other mathematicians of the 1840's. However, in the course of time all these ingredients developed into a program which turned some very vague and descriptive ideas about the 'holeyness' of solids into an impressive general theory – an algebraic context within which these ideas can be formulated independently of any intuitive notions. There are many famous names associated with this program. One of them was the Italian mathematician **Enrico Betti**, who introduced numbers, known as *Betti numbers*, which turn out to be a generalization of the *Kirchhoff number* $(E - V + 1)$.

But the person who made the greatest advances, in a series of papers published around 1895, was the French mathematician **Henri Poincaré**. He formulated everything in terms of multi-dimensional objects (*complexes*), built out of what he called *simplexes*, and he showed how the rules by which they are fitted together can be described by means of matrices. He also showed how the 'holeyness' of complexes can be described algebraically in terms of properties of these matrices. **Veblen** (1916) gave a modern treatment of Poincaré's theory.

²²² This is compatible with *Euler's formula* if we equate the number of independent circuits to $(F - 1)$.

1847 CE Johann Benedict Listing (1806–1882, Germany). Mathematician. Started the systematic study of topology as a branch of geometry, and coined the word ‘topology’. Some topological problems are found in the works of **Euler**, **Möbius** and **Cantor**, but the subject only came into its own in 1895 with the work of **Poincaré**.

1847–1852 CE Matthew O’Brien (1814–1855, England). Mathematician. A forerunner of **Gibbs** and **Heaviside**. Introduced the modern symbols for vector multiplication.

*History of the Wave Theory of Sound*²²³

The speculation that sound is a wave phenomenon grew out of observations of water waves. The rudimentary notion of a wave is that of an oscillatory disturbance that moves away from some source and transports no discernible amount of matter over large distances of propagation.

*The possibility that sound exhibits analogous behavior was emphasized by the Greek philosopher **Chrysippos** (ca 240 BCE), by the Roman architect*

²²³ For further reading, see:

- Crighton, D.G. et al, *Modern Methods in Analytical Acoustics*, Springer Verlag: Berlin, 1992, 738 pp.
- Pierce, A.D., *Acoustics*, American Institute of Physics, 1989, 678 pp.
- Dowling, A.P. and J.E. Ffowcs Williams, *Sound and Sources of Sound*, Ellis Horwood, 1983, 321 pp.
- Lord Rayleigh, *Theory of Sound*, Vols I-II, Dover: New York, 1945.
- Morse, P.M. and K.U. Ingard, *Theoretical Acoustics*, McGraw-Hill, 1968, 927 pp.

and engineer **Vitruvius** (ca 35 BCE), and by the Roman writer **Boethius**²²⁴ (ca 475–524).

The pertinent experimental result that the air motion generated by a vibrating body (sounding a single musical note) is also vibrating at the same frequency as the body²²⁵, was inferred with reasonable conclusiveness in the early 17th century by **Marin Mersenne** (1636) and **Galileo Galilei** (1638).

Mersenne's description of the first absolute determination of the frequency of an audible tone (at 84 Hz) implies that he had already demonstrated that the frequency ratio of two vibrating strings, radiating a musical note and its octave, is as 1: 2. The perceived harmony (consonance) of two such notes would be explained if the ratio of the air oscillation frequency is also 1: 2, which in turn is consistent with the source-air motion frequency equivalence hypothesis.

The analogy with water waves was strengthened by the belief that air motion associated with musical sound is oscillatory and by the observation that sound travels with finite speed. Another matter of common knowledge was that sound bends around corners, which suggested diffraction, a phenomenon often observed in water waves. Also, **Robert Boyle's** (1660) classic experiment on the sound radiation by a ticking watch in a partially evacuated glass vessel provided evidence that air is necessary, both for the production and transmission of sound.

The apparent conflict between ray and wave theories played a major role in the history of the sister science of optics, but the theory of sound developed almost from the beginning as a wave theory.

When ray concepts were used to explain acoustic phenomena (as was done by **Reynolds** and **Rayleigh** in the 19th century), they were regarded, either explicitly or implicitly, as mathematical approximations to a well-developed wave theory.

²²⁴ Born into an aristocratic Christian family and became a consul (510). He wrote texts on geometry and arithmetic which were of poor quality but used for many centuries during a time when mathematical achievements in Europe were remarkable low. Boethius fell from favor and was imprisoned and later executed for treason and magic.

²²⁵ The history of this is intertwined with the development of the laws of vibrating strings and the physical interpretations of musical consonances, which goes back to **Pythagoras** (ca 550 BCE) and perhaps earlier. Thus, the dual nature of wave-motion in both time and frequency domains goes back all the way to the ancient Greeks.

The successful incorporation of geometrical optics into a more comprehensive wave theory had demonstrated that viable approximate models of complicated wave phenomena could be expressed in terms of ray concepts. This recognition has strongly influenced 20th century development in architectural acoustics, underwater acoustics, and noise control.

The mathematical theory of sound propagation began with **Isaac Newton** (1642–1727), whose *Principia* (1686) included a mechanical interpretation of sound as being pressure pulses transmitted through neighboring fluid particles²²⁶. Substantial progress toward the development of a viable theory of sound propagation resting on firmer mathematical and physical concepts was made in 1759–1816 by **Euler**, **d’Alembert**, **Lagrange** and **Laplace**. During this era, continuum physics, or field theory, began to receive a definite mathematical structure. The wave equation emerged in a number of contexts, including the propagation of sound in air. The theory ultimately proposed for sound in the 18th century was incomplete from many standpoints, but the modern theories of today can be regarded for the most part as refinements of that developed by Euler and his contemporaries.

The linearized equations of the acoustic field are derived directly from the general equations of fluid motion on the basis that the fluid velocity \mathbf{u} , the change of pressure p , and the change of density ρ — are all small compared to the sound velocity c , average ambient density ρ_0 , and average background pressure p_0 , respectively, such that products of the small entities can be neglected in the equations.

There are three fundamental equations relating the above entities:

(1) *Newton’s equation of motion* (conservation of the fluid linear momentum) relating the pressure gradient to the linear fluid acceleration:

$$\nabla p = -\rho_0 \frac{\partial \mathbf{u}}{\partial t};$$

(2) *the equation of continuity* (conservation of mass):

$$\rho_0 \operatorname{div} \mathbf{u} + \frac{\partial \rho}{\partial t} = 0;$$

²²⁶ The fundamental relation $\lambda f = c$ [λ = wavelength; f = frequency; c = phase velocity] appeared explicitly for the first time in Newton’s *Principia* (1686). The first measurement of the sound speed in air was evidently made by **Mersenne** (1635, 1644). The time was measured from the visual sighting of a firing of a cannon to the reception of the transient sound pulse at a known distance from the source.

(3) the equation of state, specifying the functional dependence $p = p(\rho)$, subjected to the expansion

$$p - p_0 = (\rho - \rho_0) \left(\frac{dp}{d\rho} \right)_{\rho_0} + \dots \approx (\rho - \rho_0)c^2,$$

where $c^2 = \left. \frac{dp}{d\rho} \right|_{\rho_0}$ and c the ambient velocity of sound.

Newton (1686), applying Boyle's law $p = \rho f(T)$ [isothermal process], obtained $c = \sqrt{\frac{p_0}{\rho_0}} = 290 \frac{\text{m}}{\text{sec}}$ at $T = 293^\circ\text{K}$, 15% lower than the observed value.

Laplace (1816) improved on Newton's result by correctly assuming that sound waves pass too rapidly for a significant exchange of heat to take place. For an adiabatic expansion in a perfect gas, he used $p\rho^{-\gamma} = p_0\rho_0^{-\gamma}$, which led him to

$$c^2 = \frac{dp}{d\rho} = \gamma \frac{p_0}{\rho_0} = \gamma RT$$

with $c = 343 \frac{\text{m}}{\text{sec}}$ at $T = 293^\circ\text{K}$,

$$\gamma = c_p/c_v = \text{ratio of specific heats,}$$

and

$$R = \text{universal gas constant.}$$

[Clearly, the theoretical prediction of the speed of sound in liquids is more difficult than in gases. For example, c in sea water depends on the pressure, salinity, water temperature and the amount of dissolved and suspended gas.]

The above equations then imply the approximate relations

$$p = \rho c^2 + \text{const.},$$

$$k \operatorname{div} \mathbf{u} + \frac{\partial p}{\partial t} = 0,$$

where $k = \rho_0 c^2$ is the *incompressibility*. The combination of the conservation laws for mass and momentum leads to the wave equation

$$\nabla^2 p = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2}$$

for the acoustic pressure changes. The further assumption $\mathbf{u} = \operatorname{grad} \psi$, implies that the fluid velocity \mathbf{u} also obeys the same wave equation. It then follows that all field entities are expressible in terms of the potential ψ :

$$p - p_0 = -\rho_0 \frac{\partial \psi}{\partial t};$$

$$\rho - \rho_0 = -\frac{\rho_0}{c^2} \frac{\partial \psi}{\partial t}; \quad \mathbf{u} = \nabla \psi.$$

The wave equation for ψ is

$$\nabla^2 \psi = \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2}.$$

For one-dimensional motion,

$$\psi = \psi(x - ct); \quad u = \psi'$$

implies at once the relations $p = \rho_0 c u$; $\rho = \rho_0 \frac{1}{c} u$. Certain entities formed of the basic field elements $\{p, u, c, \rho_0\}$ are of use in acoustic engineering:

$$Z = \rho_0 c \equiv \sqrt{\rho_0 k} \quad (\text{impedance});$$

$$W = \frac{1}{2} \rho_0 |u|^2 + \frac{1}{2k} |p|^2 = \rho_0 |u|^2 = \frac{|p|^2}{\rho_0 c^2}$$

(wave energy density = fluid momentum flux);

$$I = pu = Wc$$

(sound intensity = rate at which acoustic energy crosses a unit area per unit time).

The application of the Fourier transform to the pressure wave equation yields the Helmholtz equation (1860):

$$\nabla^2 p + \frac{\omega^2}{c^2} p = 0,$$

where ω is the angular frequency of the harmonic Fourier component.

It is of interest to note that **Euler**, in his “Continuation of the Researches on the Propagation of Sound” (1759, 1766), already derived the Helmholtz spectral wave equation for the particle displacement (or velocity).

The solution of the Helmholtz equation for a symmetrical point-source yields the well-known result that the sound intensity falls off as the square of the distance from the source in a free open space. For sources of large area, the approximation does not hold and the sound intensity may at first fall off proportionally to the first power of the distance. Finally, in enclosed regions the sound intensity may decrease very slowly, or not at all, with distance.

Pressure is measured in units of Pascal, denoted $Pa = \frac{\text{Newton}}{\text{m}^2} = 10 \frac{\text{dyn}}{\text{cm}^2}$ [Newton = 10^5 dyn, Joule = Newton \times meter; Watt = Newton \times meter/sec].

Another unit is the $\text{bar} = 10^5 \text{ Newton/m}^2 = 10^6 \text{ dyn/cm}^2 \approx \text{Kg/cm}^2$;
 $1\mu\text{bar} = 10^{-6} \text{ bar} = \text{dyn/cm}^2$. Atmospheric pressure $\approx 10^5 \text{ Pa} \approx \text{Kg/cm}^2$.

1847–1856 CE Jean Frederic Frenet (1816–1888, France). Mathematician. Contributed to differential geometry of curves and surfaces. Introduced the so-called *Frenet-Serret*²²⁷ *formulae* for the moving-trihedral on a space curve. He was a man of wide erudition and a classical scholar.

Frenet was born at Perigueux and graduated from the École Normale Superior (1840). He was a professor at Toulouse and Lyons.

1847–1861 CE Ignaz Philipp Semmelweis (1818–1865, Hungary). Obstetrician. Pioneer of antiseptis²²⁸. Proved (1847–1849) that *puerperal fever* (childbed fever) is brought to the woman in labor by the hands and instruments of examining physicians and can be eliminated through a thorough cleansing, in a solution of water with chloride of lime, of the hands, instruments, and other items brought in contact with the patient. Published (1861) *Die Aetiologie, der Begrift und die Prophylaxis des Kindbettfiebers*.

Semmelweis was born in Buda to Jewish parents and was educated at the Universities of Pest and Vienna, graduating M.D. in 1844. At the time when he was appointed assistant professor in a maternity ward, the mortality rate from puerperal fever stood at about 20 percent. His antiseptic measures caused this rate to drop to 1.2 percent by May 1847. His superior, Johann Klein, apparently blinded by jealousy and vanity, and supported by other reactionary teachers, drove Semmelweis from Vienna (1849).

Fortunately, in the following year Semmelweis was appointed obstetric physician at Pest in the maternity department, then as terribly afflicted as Klein's clinic had been. In the course of his six years of tenure there he succeeded, by antiseptic methods, in reducing the mortality rate to 0.85 percent. However, constant conflicts with his uncooperating superiors brought

²²⁷ **Joseph Alfred Serret** (1819–1885, France). Mathematician. Graduated from the Ecole Polytechnique (1840). Professor of celestial mechanics at College de France (1861); Professor of Mathematics at Sorbonne (1863). Succeeded Poinot in the Academie des Sciences (1860).

²²⁸ In 1854, **Heinrich Schröder** and **Theodor von Dusch** showed that bacteria could be removed from air by filtering through cotton-wool. In 1867, **Joseph Lister** (1827–1912, England) reported his method of antiseptic surgery [son of Joseph Jackson Lister (1786–1869)].

him within the gates of an asylum (1865). He brought with him into this retreat an infected dissection wound which caused his death — a victim of the very disease for the relief of which he had already sacrificed health and fortune.

1847–1894 CE Hermann Ludwig Ferdinand von Helmholtz (1821–1894, Germany). One of the foremost scientists of the 19th century. Surgeon, physiologist, physicist, mathematician, chemist, musical scientist and philosopher. Helmholtz was among the last of the universalists: his research spanned almost the entire gamut of science.

In one of the epoch-making papers of the century, he formulated in 1847 the universal law of conservation of energy. Presented (1858) the first mathematical account of *rotational* fluid flow, introducing the important concepts of *vorticity*, *circulation*, *vortex flow*²²⁹ and *vortex lines*. In 1860, Helmholtz

²²⁹ Circulatory flow that is irrotational everywhere (except possibly at $r = 0$) is possible and is known as *circulatory flow without rotation*. In this case if the fluid is also incompressible and the flow *stationary* the velocity field has to satisfy both the matter conservation ($\text{div}(\mathbf{V}) = 0$) and irrotationality ($\boldsymbol{\Omega} = \text{curl } \mathbf{V} = 0$) conditions. The simplest solution of this class exhibiting circulatory flow about $r = 0$ has

$$\mathbf{V} = u_{\theta}(r)\mathbf{e}_{\theta},$$

while irrotationality requires

$$\boldsymbol{\Omega} = \frac{\partial(ru_{\theta})}{r\partial r} = 0.$$

It therefore follows that

$$ru_{\theta} = K = \text{constant}$$

(which is the law of conservation of angular momentum in disguise; the fluid angular-momentum density is $J = \rho u_{\theta} r$). Thus

$$\mathbf{V} = K \frac{\mathbf{e}_{\theta}}{r}$$

representing irrotational motion except at the point $r = 0$, where the vorticity $\boldsymbol{\Omega}$ and the velocity become infinite (this is obviously an idealization of actual such flows). The circulation along a streamline $r = \text{const.}$ is

$$\Gamma = \oint \mathbf{V} \cdot d\boldsymbol{\ell} = 2\pi K$$

and the motion is known as *vortex flow*. It plays an important role in aerodynamics. On the basis of experimental evidence and the theory of viscous flow, one can assume that there is a fluid core or nucleus surrounding the center of

developed the mathematical theory of Huygens' principle for 'monochromatic' steady-state scalar waves. He also showed that an arbitrary continuously differentiable vector-field can be represented at each point as a superposition of the gradient of a scalar potential and a curl of a vector potential.²³⁰

Helmholtz made a great contribution to our understanding of thermodynamics; he was first to apply minimum principles to thermodynamics, and showed that for reversible processes, the role of the *action* was played by the "*Helmholtz free energy*", F .

In 1854 Helmholtz seized upon the problem of the sun's luminosity. Previously, **Kant** had calculated that if the sun's light came from ordinary combustion, it would have burned up in only 3000 years. Helmholtz then argued that the tremendous weight of the sun's outer layers, pressing radially inward, should cause the sun to gradually contract: Consequently, its interior gases will become compressed, and heat up. Hence gravitational contraction causes the sun's gases to become hot enough to radiate energy into space. He was thus able to boost the theoretical age of the sun to some 20 million years. This in turn meant that the sun extended beyond the earth's orbit only 20 millions years ago, to which geologists could not agree on the basis of the earth's present surface features. **Kelvin** supported and 'improved' Helmholtz's theory and it is known as *Helmholtz-Kelvin contraction*.

In other fields of science, Helmholtz contributed to the subjects of: fermentation, animal heat and electricity, muscular contraction, velocity of nerve

the flow and that the core rotates approximately like a solid body. Within the core we have circulatory flow with constant velocity and outside the core we have circulatory flow without rotation. Inside the core

$$u_{\theta} \sim r$$

while outside

$$u_{\theta} \sim \frac{1}{r}.$$

Such a combination is known as an *eddy* or simply a *vortex*. The central core is called the vortex core. The *tornado* and *water spout* (or even the common *bathub vortex*) are examples of such a flow. The *stability* of the vortex is determined by its Reynolds' number.

If an eddy occurs in a fluid that is otherwise undisturbed, the spatial location of the eddy remains unaltered. However, if a uniform stream is superposed on it, it will *move* with the stream. Such a vortex is known as a *free vortex*.

²³⁰ This theorem is now recognized as a special case of a result from Cartan's exterior calculus in an arbitrary, n -dimensional manifold. The more general result relates to algebraic topology through the *de Rham cohomology*.

impulses²³¹, invention of the ophthalmometer, physiological optics, color vision, physiological acoustics and meteorological physics.

From 1869 to 1871 Helmholtz involved himself in the verification of Maxwell's predictions concerning electromagnetic waves. He entrusted the subject into the hands of his favorite pupil, **Heinrich Hertz**, and the latter finally gave an experimental verification of their existence and velocity.

Helmholtz was born in Potsdam, near Berlin. His father was a high school teacher and his mother was a lineal descendant of the Quaker **William Penn** (founder of the state of Pennsylvania).

As his parents were poor and could not afford to allow him to pursue a purely scientific career, he became a surgeon in the Prussian army. He lived in Berlin from 1842 to 1849, when he became a professor of physiology in Königsberg. In 1855 he removed to assume the chair of physiology in Bonn. In 1858 he became professor of physiology at Heidelberg, and in 1871 he was called to occupy the chair of physics in Berlin.

Helmholtz married twice and had 4 children. He was a man of simple but refined tastes, noble carriage and somewhat austere manner. His life, from first to last, was one of devotion to science.

1848 CE A year of revolutions in almost every European country. It was the natural climax of a process of reaction and revolt which began after the defeat of Napoleon at Waterloo in 1815. Thereafter, Europe entered a period of instability, characterized by a long series of upheavals. The revolution of 1848 was the culmination of the political, economical and social unrest of the time — of the struggle between the aristocracy and the middle classes, the rapid increase of population from 180 million in 1800 to 266 million in 1850, the fact that more and more people now lived in cities, the conflict between the bourgeoisie and the rising proletariat, and the movements for national liberation and reunion. And it confounded all the protagonists, compelling a reappraisal of ideas and a realignment of forces.

In some sense, the French Revolution and its sequel in Napoleonic imperialism, disrupted the historic continuity of European society and shattered most of its traditions.

All the significant problems of the period arose out of these events. This break in continuity engendered a quest for new patterns of interpretation — nationalism, socialism, vast philosophical systems like those of **Marx** and **Hegel**, new conceptions of historical, scientific, literary and artistic ideas.

²³¹ He actually measured the speed of nerve impulses (1852).

Table 4.4 TIMELINE OF THE INDUSTRIAL REVOLUTION, 1770–1848

- ca 1770 — Consolidation of the steam engine by **James Watt**
- 1775–1783 — American Revolution
- 1780 — Industrial Revolution under way
- 1789–1794 — The French Revolution
- 1799–1815 — Reign of Napoleon
- 1800–1850 — Romanticism in literature and the arts
- 1815 — The Congress of Vienna and the congress system of European diplomacy
- 1820 — Revolutions in Greece and Spain
- 1830 — Rise of liberalism and nationalism
- 1830, 1848 — Periods of revolution in Europe
- 1832 — Parliamentary reform in Great Britain
- 1848 — **Karl Marx's** 'Communist Manifesto'.

Europe's search for stability after 1815 was marked by a contest between the forces of the past and the forces of the future. For a while it seemed as though the traditional agencies of power — the monarchs, the landed aristocracy and the Church — might once again resume full control. But potent new forces were ready to oppose relapse into the past. With the quickening of industrialization, there was now not only a middle class of growing size and significance but a wholly new class, the urban proletariat. Each class had its own political and economical philosophy — liberalism and socialism respectively — which stood opposed to each other as well as to the traditional conservatism of the old order.

Nationalism as an awareness of belonging to a particular nationality was nothing new. What was new was the *intensity* that this awareness now assumed: for the mass of the people, nationalism became their most ardent emotion, and national unification or independence their most cherished aim.

The Vienna settlement (1815) ignored the stirrings of nationalism and the hopes for democracy that had been awakened by the French Revolution. It was mainly interested in peace and order and the restoration of conditions as they

were before the French Revolution²³². Indeed, there was no war among the great European powers for 40 years, and no war of world-wide dimensions for a whole century. The Triple Alliance of Austria, Russia and Prussia guaranteed to maintain the territorial *status quo* in Europe and the existing form of government in every European country, i.e. aiding legitimate governments against revolutions.

A first wave of reaction that followed the peace settlements of 1815, manifested itself in the first wave of revolutions (1820–1829) in Spain, Portugal, Italy, Greece and Russia. The second wave of Revolutions swept France, Belgium and Poland during 1830–1833. The third wave (1848–1849) lasted for over a year and affected most of Europe with the exception of England and Russia. In Italy, Germany, Austria, and Hungary, the fundamental grievance was still the lack of national freedom and unity. In Western Europe the chief aim of revolutions was the extension of political power beyond the upper middle class. With the revolutions of 1848, socialism for the first time became an issue of modern politics.

In addition, severe economic crises particularly affected the lower classes: everywhere the small artisan was fighting against the competition of large-scale industry, which threatened to deprive him of his livelihood. At the same time, the industrial workers in the new factories were eeking out a miserable existence on a minimum wage. There were also periodic upheavals in agriculture, primarily as a result of crop failures.

The revolutions of 1848 had failed everywhere due to weaknesses in the revolutionary camp (lack of widespread popular support, indecision among their leaders and the lack of well-defined programs) and the continued strength of the forces of reaction. The burden of the revolution fell on the workers whereas the middle class, in most countries, did not really want a revolution. It preferred to achieve its aims through reform, as had been done in England. There was no attempt to coordinate the revolutions in different countries, although the forces of reaction worked together.

Two forces emerged from the revolutions that henceforth were to dominate the history of Europe — nationalism and socialism. These now became, respectively, the main issues in the struggle of nation against nation and class against class.

²³² In Spain and Naples the returning Bourbons abolished the liberal reforms that had been granted in 1812. In the Papal States, Pope Pius VII got rid of the French legal reforms, re-established the Jesuits, put the Jews back into the ghettos, and forbade vaccination against smallpox!

In Piedmont, Victor Emmanuel I had the French botanical gardens torn up by the roots and the French furniture thrown out of the windows of his palace!



<http://www.springer.com/978-3-540-68831-0>

Historical Encyclopedia of Natural and Mathematical
Sciences

Ben-Menahem, A.

2009, CLXII, 5983 p. In 6 volumes, not available
separately., Hardcover

ISBN: 978-3-540-68831-0