Chapter 2

# MATHEMATICAL BACKGROUND

## *2.1 Geometry*

The formalization of geometry began with the ancient Greeks. They took what had been an *ad hoc* collection of surveying and measuring tools and rebuilt them on top of the bedrock of logic. A remarkable journey began then and continues to the present day. The story is brilliantly told by Greenberg [62] and summarized concisely by Stewart [130, Chapter 5]. Other valuable presentations of Euclidean and non-Euclidean geometry are given by Bonola [14], Coxeter [25], Faber [51], Kay [97], and Martin [109].

In his monumental *Elements*, Euclid attempts to reduce the study of geometry to a minimal number of required assumptions from which all other true statements may be derived. He arrives at five postulates, primitive truths that must be accepted without proof, with the rest of plane geometry following as a reward. Over the years, some of the postulates (particularly the fifth) have drifted to alternate, logically equivalent forms. One statement of Euclidean geometry, adapted from Greenberg [62, Page 14], is as follows:

I. Any two distinct points lie on a uniquely determined line.

II. A segment $AB$ may always be extended by a new segment $BE$ congruent to a given segment $CD$.

III. Given points $O$ and $A$, there exists a circle centered at $O$ and having segment $OA$ as a radius.

IV. All right angles are congruent.

V. (*Playfair's postulate*) Given a line $l$ and a point $P$ not on $l$, there exists a unique line $m$ through $P$ that is parallel to $l$.

The development of a logical system such as Euclid's geometry is a process of abstraction and distillation. Euclid presented his five postulates as the basis for all of geometry, in the sense that any other true statement in geometry could be derived from these five using accepted rules of logical deduction. Today, we know that Euclid's postulates are incomplete — they are an insufficient foundation upon which to build all that he wanted to be true. Many mathematicians have since provided revised postulates that preserve the spirit of Euclid's geometry and hold up under the careful scrutiny of modern mathematics. In all cases the core idea remains to distill all possible truths down to a minimal set of intuitive assertions.

There are two other senses in which this distillation must occur in order to make the logical foundation of geometry self-contained. First, the *objects* of discourse must be reduced to a suitable primitive set. The postulates mention only points, lines, segments, circles, and angles. No mention is made of polygons, parabolas, or a multitude of other geometric objects, because all such objects can be defined in terms of the five mentioned in the postulates. Even segments, circles, and angles can be defined in terms of points and lines. Euclid attempts to take this process to the limit, providing definitions for points and lines. However, his definitions are somewhat enigmatic. Today, we know that just as all truth must eventually bottom out in a set of primitive postulates, all identity must reduce to a set of primitive objects. So we reduce plane geometry to two sorts of objects: points and lines. These objects require no definition; as Hilbert famously remarked, geometry should be equally valid if it were phrased in terms of tables, chairs, and beer mugs. The behaviour of these abstract points and lines is determined by the postulates. We keep the names as an evocative reminder of the origins of these objects.

The other chain of definition concerns the relationships between objects. The postulates mention relationships like "lie on," "congruent," and "parallel." Again, the chain of definition must bottom out with some primitive set of relationships from which all others can be constructed. In modern presentations of Euclidean geometry such as Hilbert's [62, Chapter 3], three relationships are given as primitive: *incidence*, *congruence,* and *betweenness*. Incidence determines which points lie on which lines. Congruence determines when two segments or two angles have "the same shape." Betweenness is implicit in the definition of objects like segments (the segment $AB$ is $A$ and $B$ together with the set of points $C$ such that $C$ is between $A$ and $B$). Again, these relationships do not have any *a priori* definitions; their behaviour is specified and constrained through their use in the postulates.

We are left then with geometry as a purely logical system (a *first-order language*, in mathematical logic [81, Section 4.2]). When establishing the validity of a statement in geometry, any connection to its empirical roots is irrelevant. We call this system *Euclidean geometry*, or sometimes *parabolic geometry*.

In a sense though, geometry is still “about” objects like points and lines. Geometry can be tied back to a concrete universe of points and lines through an *interpretation*. An interpretation of Euclidean geometry is a translation of the abstract points and lines into well-defined sets, and a translation of the incidence, congruence, and betweenness relations into well-defined relations on those sets. The postulates of geometry then become statements in the mathematical world of the interpretation. An interpretation is called a *model* of geometry when all the postulates are true statements.

The familiar Cartesian plane, with points interpreted as ordered pairs of real numbers, is a model of Euclidean geometry. But it is a mistake to say that the Cartesian plane *is* Euclidean geometry. Other inequivalent models are possible for the postulates given above. It was only in the nineteenth century, when the primacy of Euclidean geometry was finally called into question, that mathematicians worked to rule out these alternate models and make the logical framework of geometry match up with the intuition it sought to formalize. Several systems (such as Hilbert’s) emerged that were *categorical*: every model of the system is isomorphic to the Cartesian plane. In such categorical systems, it is once again safe to picture Euclidean geometry as Euclid did, in terms of the intuitive notions of points and lines.

### *2.1.1 Hyperbolic geometry*

The fifth postulate, the so-called “parallel postulate,” is the source of one of the greatest controversies in the history of science, and ultimately led to one of its greatest revolutions.

In any logical system, the postulates (also called the *axioms*) should be obvious, requiring only a minimal investment of credulity. From the start, however, the parallel postulate was considered much too complicated, a lumbering beast compared to the other four. Euclid himself held out as long as possible, finally introducing the parallel postulate in order to prove his twenty-ninth proposition.

For centuries, mathematicians struggled with the parallel postulate. They sought either to replace it with a simpler, less contentious axiom, or better yet to establish it as a consequence of the first four postulates. Neither approach proved successful; no alternate postulate was found that was uncontroversial, and any attempts to formulate a proof either led to a dead end or involved a hidden assumption, itself equivalent to the parallel postulate.

All of these efforts were based on one fundamental hidden assumption: that the only possible geometry was that of Euclid. The possibility that an unfamiliar but perfectly valid geometry could exist without the parallel postulate was unthinkable. Literally so, according to Kant, who in his *Critique of Pure Reason* declared that Euclidean geometry was not merely a fact of the physical universe, but inherent in the very nature of thought [62, Page 182].

Finally, in the nineteenth century, a breakthrough was made by three mathematicians: Bolyai, Gauss, and Lobachevski. They separately realized that the parallel postulate was in fact *independent* of the rest of Euclidean geometry, that it could be neither proven nor disproven from the other axioms. Each of them considered an alternate logical system based on a modified parallel postulate in which multiple lines, all parallel to $l$, could pass through point $P$. This new geometry appeared totally self-consistent, and indeed was later proven to be so by Beltrami.[1] Paulos [117] likens the consistency of non-Euclidean geometry to the surprising but plausible incongruity that makes riddles funny – the riddle in this case being "What satisfies the first four axioms of Euclid?"

Today, we refer to the non-Euclidean geometry of Bolyai, Gauss, and Lobachevsky as *hyperbolic geometry*, the study of points and lines in the *hyperbolic plane*. Hyperbolic geometry is based on the following alternate version of the parallel postulate:

V. Given a line $l$ and a point $P$ not on $l$, there exist at least two lines $m_1$ and $m_2$ through $P$ that are parallel to $l$.

In Euclidean ornamental designs, parallel lines can play an important role. To thicken a mathematical line $l$ into a band of constant width $w$, we can simply take the region bounded by the two parallels of distance $w/2$ from $l$. This approach presents a problem in hyperbolic geometry, where these parallels are no longer uniquely defined. On the other hand, parallelism is not the defining quality of a thickened line, merely a convenient Euclidean equivalence. What we are really after are

[1] Beltrami's proof hinged upon exhibiting a model of non-Euclidean geometry in the Euclidean plane. Any inconsistency in the logical structure of non-Euclidean geometry could then be interpreted as an inconsistency in Euclidean geometry, which we are assuming to be consistent. This sort of *relative consistency* is about the best one could hope for in a proof of the validity of any geometry.

the loci of points of constant perpendicular distance $w/2$ from $l$. These are called *equidistant curves* or *hypercycles*, and they are always uniquely defined. In the Euclidean plane, equidistant curves are just parallel lines. In the hyperbolic plane, they are curved paths that follow a given line.

There are several different Euclidean models of hyperbolic geometry; all are useful in different contexts. Each has its own coordinate system. Hausmann *et al.* [78] give formulae for converting between points in the three models.

In the *Poincaré model*, the points are points in the interior of the Euclidean unit disk, and the lines are circular arcs that cut the boundary of the disk at right angles (we extend this set to include diameters of the disk). The Poincaré model is *conformal*: the angle between any two hyperbolic lines is accurately reflected by the Euclidean angle between the two circular arcs [2] that represent them. The Poincaré model is therefore a good choice for drawing Euclidean representations of hyperbolic patterns, because in some sense it does the best job of preserving the “shapes” of hyperbolic figures. It also happens to be particularly well-suited to drawing equidistant curves; in the Poincaré model, every equidistant curve can be represented by a circular arc that does not cut the unit disk at right angles.

The points of the *Klein model* are again the points in the interior of the Euclidean unit disk, but hyperbolic lines are interpreted as chords of the unit disk, including diameters. The Klein model is *projective*: straight hyperbolic lines are mapped to straight Euclidean lines. This fact makes the Klein model useful for certain computations. For example, the question of whether a point is inside a hyperbolic polygon can be answered by interpreting it through the Klein model as a Euclidean point-in-polygon test.

The *Minkowski model* [40, 51] requires that we move to three dimensional Euclidean space. Here, the points of the hyperbolic plane are represented by one sheet of the hyperboloid $x^2 + y^2 - z^2 = -1$, and lines are the intersections of Euclidean planes through the origin with the hyperboloid. The advantage of The Minkowski model is that rigid motions (see Section 2.2 for more on rigid motions) can be represented by three dimensional linear transforms. Long sequences of motions can therefore be concatenated via multiplication, as they can in the Euclidean plane. Our software implementations of hyperbolic geometry are based primarily on the Minkowski model, with points

---

[2] The angle between two arcs is measured as the angle between their tangents at the point of intersection.

converted to the Poincaré model for output.

Note that although there are several models for hyperbolic geometry (including others not discussed here), it is still categorical. The Poincaré, Klein, and Minkowski models are all isomorphic [62, Page 236].

### 2.1.2 *Elliptic geometry*

Given a line $l$ and a point $P$ not on $l$, we have covered the cases where exactly one line through $P$ is parallel to $l$ (Euclidean geometry) and where several lines are parallel (hyperbolic geometry). One final case remains to be explored:

V. Given a line $l$ and a point $P$ not on $l$, every line through $P$ intersects $l$.

Once again, this choice of postulate leads to a self-consistent geometry, called *elliptic geometry*. In elliptic geometry, parallel lines simply do not exist.

A first attempt at modeling elliptic geometry would be to let the points be the surface of a three dimensional Euclidean sphere. Lines are interpreted as great circles on the sphere. Since any two distinct great circles intersect, the elliptic parallel property holds. This model is invalid, however, because Euclid's first postulate fails. Antipodal points lie on an infinite number of great circles.

A strange but simple modification to the spherical interpretation can make it into a true model of elliptic geometry. A point is interpreted as a *pair* of antipodal points on the sphere. Lines are still great circles. The identification of a point with its antipodal counterpart fixes the problem with the first postulate, because no elliptic "point" is now more than a quarter of the way around the circle from any other, and the great circle joining those points is uniquely defined.

Despite this antipodal identification, the elliptic plane can still be drawn as a sphere, with the understanding that half of the drawing is redundant. Any elliptic figure will be drawn twice in this representation, the two copies opposite one another on the sphere. Note also that the equidistant curves on the sphere are simply non-great circles.

### 2.1.3 *Absolute geometry*

The parallel postulate is independent from the other four, which allows us to choose any of the three alternatives given above and obtain a consistent geometry. But what happens if we choose *none* of them? In other words, let us decide to leave the behaviour of parallel lines undefined, and develop that part of geometry that does not depend on parallelism. We refer to this geometry, based only on the first four postulates, as *absolute geometry*.

Formally, this choice presents no difficulties whatsoever. We have already assumed that the first four postulates are consistent, and so they must lead to some sort of logical system. Furthermore, we already know that many Euclidean theorems still hold in absolute geometry; these are the ones whose proofs do not rely on the parallel postulate. The first twenty-eight of Euclid's propositions have this property.

In practice, the absolute plane is somewhat challenging to work with. As always, in order to visualize the logical system represented by absolute geometry, we need a model. Such models are easy to come by, because any model of parabolic, hyperbolic, or elliptic geometry is automatically a model of absolute geometry! Of course, those models do not tell the whole story (or rather, they tell more than the whole story), because in each case parallels have some specific behaviour. This behaviour does not invalidate the model, but it imposes additional structure that can be misleading. It is perhaps easier to imagine absolute geometry as a purely formal system, one that contains all the constructions that are common to parabolic, hyperbolic, and elliptic geometry.

The model of elliptic geometry presented above can be somewhat difficult to visualize and manipulate. In some ways, it would be desirable to work directly with the sphere with no identification of antipodal points. From there, perhaps an absolute geometry could be developed that unifies the Euclidean plane, the hyperbolic plane, and the sphere in a natural way.

Unfortunately, as we have seen, the native geometry on the sphere violates Euclid's first postulate. However, it turns out that by giving a slightly revised set of axioms, we can in fact develop a consistent geometry modeled by the Euclidean sphere without the identification of antipodal points. This geometry is called *spherical geometry*, or sometimes *double elliptic geometry*. Moving from elliptic to spherical geometry requires some reworking of Euclid's postulates, but is justified by the convenience of a far more intuitive model.

Kay [97] develops an axiomatic system for spherical geometry. The trick is to start with *ruler and protractor postulates*, axioms that provide formal measures of distance and angle. A (possibly infinite) real number $\alpha$ is then defined as the supremum of all possible distances between points. On the sphere, $\alpha$ is half the circumference; in the Euclidean and hyperbolic planes, $\alpha$ is infinite. Kay then insinuates $\alpha$ into his axioms, using it to do the bookkeeping necessary to avoid problematic situations. For instance, his version of Euclid's first postulate is as follows:

I. Any two points $P$ and $Q$ lie on at least one line; when the distance from $P$ to $Q$ is less than $\alpha$, the line is unique.

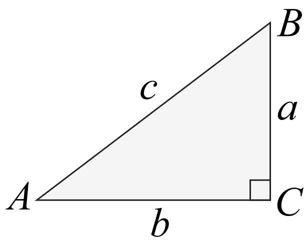
Kay's presentation carefully postpones any assumption on parallelism until the final axiom. As a result, we can consider the geometry formed by all the axioms except the last one. This is a form of absolute geometry that can be specialized into parabolic, hyperbolic, and spherical (as opposed to elliptic) geometry.

In the absence of any single model that exactly captures its features, one may wonder how absolute geometry can be made practical. We do know that any theorem of absolute geometry will automatically hold in parabolic, hyperbolic, and spherical geometry, since formally they are all just special cases. By interpreting absolute geometry in various different ways, we can then view that theorem as a true statement about the Cartesian plane, the Poincaré disk, the sphere, or any of the other models discussed above. In effect, we can imagine an implementation of absolute geometry that is parameterized over the model. In computer science terms, the interface has no inherent representation of points or lines, but interprets the axioms of absolute geometry as a contract that will be fulfilled by any implementation. A client program can be written to that contract, and later instantiated by plugging in any valid model. This approach will be discussed in greater detail in Section 3.6.

### *2.1.4 Absolute trigonometry*

Trigonometry is the study of the relationships between parts of triangles. Triangles exist in absolute geometry and can therefore be studied using *absolute trigonometry*. Bolyai first described some of the properties of absolute triangles; his work was later expanded upon by De Tilly [14, Page 113].

Following De Tilly, we define two functions on the real numbers: $\bigcirc(x)$ and $\mathrm{E}(x)$. The function $\bigcirc(x)$, or "circle-of-$x$," is defined as the circumference of a circle of radius $x$. To obtain a definition for $\mathrm{E}(x)$, let $l$ be a line and $c$ be an equidistant curve erected at perpendicular distance $x$ from $l$. Take any finite section out of the curve, and define $\mathrm{E}(x)$ as the ratio of the length of that segment to the length of its projection onto $l$. It can be shown that this value depends only on $x$. From these




$$\bigcirc(a) = \bigcirc(c)\sin A \quad \cos A = \mathrm{E}(a)\sin B$$
$$\bigcirc(b) = \bigcirc(c)\sin B \quad \cos B = \mathrm{E}(b)\sin A$$

$$\mathrm{E}(c) = \mathrm{E}(a)\mathrm{E}(b)$$

**Figure 2.1** Trigonometric identities for a right triangle in absolute geometry. When interpreted in Euclidean geometry, the equation $\bigcirc(a) = \bigcirc(c)\sin A$ becomes "sine equals opposite over hypoteneuse," and $\cos A = \mathrm{E}(a)\sin B$ becomes $\cos A = \sin(\frac{\pi}{2} - A)$. The equation $\mathrm{E}(c) = \mathrm{E}(a)\mathrm{E}(b)$ is vacuously true (and therefore not particularly useful) in Euclidean geometry, but in hyperbolic and spherical geometry it can be seen as one possible analogue to the Pythagorean theorem.

two functions, we can also define $\mathrm{T}(x) = \bigcirc(x)/\mathrm{E}(x)$, and the three inverses $\bigcirc^{-1}(x)$, $\mathrm{E}^{-1}(x)$, and $\mathrm{T}^{-1}(x)$.

Bolyai initiated absolute trigonometry with the observation that the sines of the angles of a triangle are as the circumferences of circles with radii equal to the lengths of the opposite sides. Expressed in the notation just given, we can say that for any triangle with vertices $ABC$ and opposite edges $abc$, $\bigcirc(a)/\sin A = \bigcirc(b)/\sin B = \bigcirc(c)/\sin C$. By substituting the Euclidean definition of $\bigcirc(x)$, this relationship can be seen as a generalization of the sine law to absolute geometry. Other identities of absolute geometry that apply specifically to right triangles are summarized in Figure 2.1.

We can give formulae for $\bigcirc(x)$ and $\mathrm{E}(x)$, though their definitions must be broken down into cases. Each case corresponds to the choice of a parabolic, hyperbolic, or spherical model.

| | parabolic | hyperbolic | spherical |
|---|---|---|---|
| $\bigcirc(x)$ | $2\pi x$ | $2\pi \sinh x$ | $2\pi \sin x$ |
| $\mathrm{E}(x)$ | $1$ | $\cosh x$ | $\cos x$ |

While it is perfectly valid to analyze absolute triangles abstractly using absolute trigonometry, actual values for side lengths and vertex angles can be derived only by resorting to one of the models. This curious fact follows from the difference between the axiomatic and analytic views of geometry. Because Euclidean geometry is categorical, the usual trigonometric functions in the Cartesian plane are the only ones (up to isomorphism) that satisfy the axiomatic relationships in a Euclidean triangle. However, absolute geometry is not categorical, meaning that different, inequivalent trigonometric functions can (and do) hold under different models. We are not used to making this distinction, because in ordinary Euclidean geometry there effectively is no distinction. It can be challenging to one's intuition to visualize such functions that are well-defined formally but not analytically.

## ***2.2 Symmetry***

The mathematical tools behind a formal treatment of symmetry are relatively new, but our appreciation of symmetric patterns goes back millennia [137]. In addition to its usefulness in many branches of science, symmetry is often used to study art and ornament [127, 135]. M.C. Escher interacted with the growth of symmetry theory, creating new art based on the mathematical results that emerged during his lifetime [124].

The original conception of symmetry, as conveyed by the dictionary definition, is expressed with words such as beauty, balance, and harmony. The word was and still is used to refer to a balance of components in a whole.

The contemporary non-scientific usage of the word, as Weyl points out, refers to an object whose left and right halves correspond through reflection in a mirror [137]. Thus a human figure, or a balance scale measuring equal weights, may be said to possess symmetry.

In light of the formal definition of symmetry to come, we qualify the particular correspondence described above as "bilateral symmetry." Bilateral symmetry is certainly a familiar experience in the world around us; it is found in the shapes of most higher animals. The prevalence of bilateral symmetry can be explained in terms of the body's response to its environment. Whereas gravity dictates specialization of an animal from top to bottom and locomotion engenders differentiation between front and back, the world mandates no intrinsic preference for left or right [137, Page 27]. An animal must move just as easily to the left as to the right, resulting in equal external structure on each side. Indeed, lower life forms whose structure is not as subject to the exigencies of gravity and linear locomotion tend towards more circular or spherical body plans.

Let us regard the mirror of bilateral symmetry as a reflection through a plane in space. Saying that the mirror reconstructs half of an object from the other half is equivalent to saying that the reflection maps the entire object onto itself. We formalize the notion of symmetry by noting two properties of this reflection. It *preserves the structure of space*, just as a (flat) mirror preserves the shapes of objects, and it *maps the object onto itself*, allowing us to think of its two halves as having the same shape. By generalizing from these two properties, we will arrive at a formal definition of symmetry.

Given a mathematical space $\mathcal{S}$, we identify some important aspect of the mathematical structure of $\mathcal{S}$, and define a set of *motions* $\mathcal{M}$ to be automorphisms of $\mathcal{S}$ that preserve that structure. Then, given some figure $F \subseteq \mathcal{S}$, we can say that a motion $\sigma$ is a *symmetry* of $F$ if $\sigma(F) = F$, that is, if $\sigma$ maps the figure $F$ to itself.

This somewhat vague definition achieves rigour when we give a specific meaning to "mathematical structure." As a simple example, let $\mathcal{S}$ be the integers from 1 to $n$, and consider preserving no structure of $\mathcal{S}$ beyond the set-theoretic. Then the motions $\mathcal{M}$ are just the $n!$ permutations of the members of $\mathcal{S}$, and every $k$-element figure (subset) of $\mathcal{S}$ has $k!(n-k)!$ symmetries, each one permuting the figure internally and the remaining elements of $\mathcal{S}$ externally.

The important mathematical spaces in the present work are the parabolic, the hyperbolic, and the spherical planes. We know from Kay's presentation [97] that each of these planes has a notion of distance, defined both formally in the axioms of geometry and concretely in the models. If we let $\mathcal{S}$ be the set of points in one of these planes, then we can define the motions to be the *isometries* of $\mathcal{S}$: the automorphisms that preserve distance. This point of view leads to the most important and most common definition of symmetry: a symmetry of a set $F$ is an isometry that maps $F$ to itself. We will also sometimes use the term *rigid motion* in the place of isometry; isometries are rigid in the sense that they do not distort shape. In the Euclidean plane, the symmetries of a figure are easily visualized by tracing the figure on a transparent sheet and moving that sheet around the plane, possibly flipping it over, so that that original figure and its tracing line up perfectly [68, Page 28].

### *2.2.1 Symmetry groups*

For a given $F \subseteq \mathcal{S}$, let $\Sigma(F)$ be the set of all motions that are symmetries of $F$. This set has a natural group structure through composition of automorphisms. The set $\Sigma(F)$ is therefore called the *symmetry group* of $F$.

The *orbit* of a point $p \in \mathcal{F}$ under a symmetry group $G$ is the set $\{\sigma(p)|\sigma \in G\}$. When $\mathcal{S}$ is equipped with a measure of distance, we say that the symmetry group $G$ is *discrete* if for every point $p$, the orbit of $p$ does not get arbitrarily close to $p$. More precisely, if $d(p, q)$ is the distance between points $p$ and $q$ in $\mathcal{S}$, then $G$ is discrete if for all $p$, $\inf\{d(p, \sigma(p))|\sigma \in G, \sigma(p) \neq p\} > 0$. A circle is an example of a figure with non-discrete symmetry; every point on the circle is a limit point of its orbit. In this work, we restrict ourselves to discrete symmetry groups, a technical but important point that simplifies the classification of the groups we will use. There exist ornamental designs that can be profitably analyzed using non-discrete symmetry groups, but such designs will not arise here.

Symmetry is a measure of redundancy in a figure, and so we ought to be able to use our understanding of the symmetries of a figure to factor out the redundancy. The result would be a minimal, sufficient set of non-redundant information that, together with the symmetries, completely describe the original figure. For any discrete symmetry group, this set exists and is called the group's *fundamental region* or *fundamental domain*. One possible definition, given by Grünbaum and Shephard [68, Section 1.6], says that $U \subseteq \mathcal{S}$ is a fundamental region of symmetry group $G$ if the following conditions hold:

(a) $U$ is a connected set with non-empty interior.

(b) No two points in $U$ have the same orbit under $G$ (or equivalently, for all $p, q \in U$, there does not exist a $\sigma \in G$ such that $\sigma(p) = q$).

(c) $U$ is as big as possible; that is, no other set satisfying (a) and (b) contains $U$ as a proper subset.

The first condition ensures that the fundamental region is relatively simple topologically. The other two guarantee that the region has "exactly enough" information — condition (b) rules out redundancy and condition (c) forces every orbit to be represented by some point in $U$.

It is important from an algorithmic standpoint to understand a symmetry group's fundamental region. The fundamental region contains a single, non-redundant copy of the information in a symmetric figure. Therefore, in order to draw the figure as a whole, it suffices to draw all images of the fundamental region under the motions of the symmetry group. The drawing process can be seen as a replication algorithm that determines the set of motions to apply, and a subroutine that draws a


**Figure 2.2** Figures in the Euclidean plane with $c_n$ or $d_n$ symmetry. Each figure is labeled with its symmetry group.

transformed copy of the fundamental region.

### *2.2.2 Some important symmetry groups*

The discrete symmetry groups in the Euclidean plane are well understood. Up to isomorphism, only a limited collection of discrete groups can be symmetry groups of Euclidean figures. They can be classified according to the nature of the subgroup of the symmetry group consisting of just translations:

- If the translational subgroup is trivial, then the symmetry group must be finite, isomorphic either to $c_n$, the cyclic group of order $n$, or $d_n$, the dihedral group of order $2n$. The group $c_n$ is the symmetry group of an $n$-armed swastika, and $d_n$ is the group of a regular $n$-gon. These possibilities were enumerated by Leonardo Da Vinci, and the completeness of the enumeration is sometimes called “Leonardo’s Theorem” [109, Section 30.1]. Examples of figures with $c_n$ and $d_n$ symmetry are shown in Figure 2.2.

- If all the translations are parallel, then the symmetry group must be one of the seven *frieze groups*. Friezes are decorations executed in bands or strips, and the frieze groups are particularly well suited to their study. A *frieze pattern* is a figure with a frieze group as its symmetry group.

- The remaining case is when the translational subgroup contains translations in two non-parallel directions. This category consists of the seventeen *wallpaper groups*. A *wallpaper*

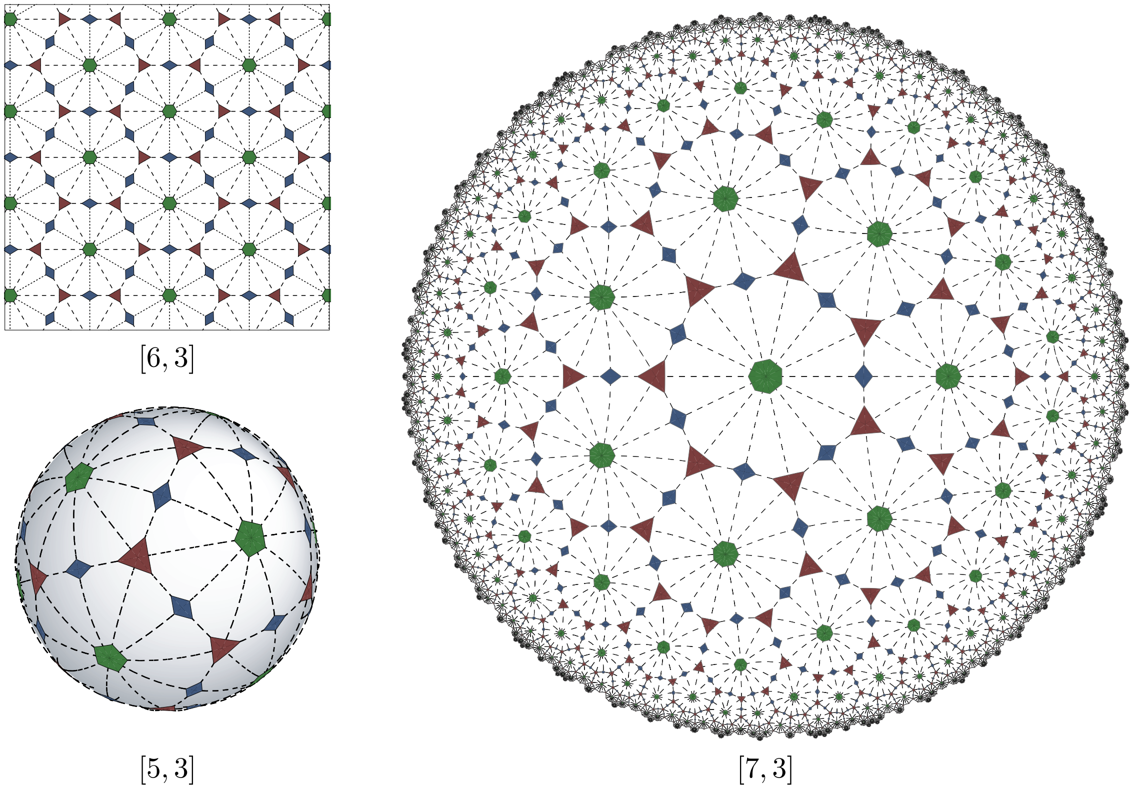



**Figure 2.3** Examples of symmetry groups of the form $[p, q]$. Each example visualizes the lines of reflection (shown as dotted lines) and centers of rotation of the symmetry group. The green, red, and blue forms represent centers of $p$-fold, $q$-fold, and twofold rotation, respectively.

*pattern* (sometimes called a *periodic pattern*) is a figure with a wallpaper group (or a *periodic group*) as its symmetry group. The wallpaper patterns are "all-over" patterns in the sense that the pattern repeats in every direction, not just in one distinguished direction as is the case with frieze patterns. A wallpaper pattern necessarily has a bounded fundamental region.

For images of frieze and wallpaper patterns, see for example Grünbaum and Shephard [68, Section 1.3], Beyer [11] (a lovely presentation in the context of quilt design), or Shubnikov and Koptsik [127]. Washburn and Crowe [135] give many examples and include a flowchart-based technique for classifying a given pattern.

The groups $c_n$ and $d_n$ can also occur as the symmetries of figures in the hyperbolic plane and on the sphere, but the frieze and wallpaper patterns are very much tied to Euclidean geometry. The linear independence implied by translational symmetry in two non-parallel directions only makes sense in an affine space; the Euclidean plane is affine, whereas the sphere and hyperbolic plane are not.

A different set of all-over patterns leads to a family of symmetry groups that span Euclidean and non-Euclidean geometry. For every $p \geq 2$ and $q \geq 2$, there is a group $[p, q]$, which can be seen as the symmetry group of a tiling of the plane by regular $p$-gons meeting $q$ around every vertex [26, Chapter 5]. These regular tilings are discussed in greater detail in Section 2.3.1.

A simple calculation shows that each group $[p, q]$ is tied to one of the three planar geometries. In particular, $[p, q]$ is a symmetry group of the Euclidean, hyperbolic, or spherical plane when $1/p+1/q$ is respectively equal to, less than, or greater than $1/2$. Note that $[p, q]$ is isomorphic to $[q, p]$, even though the regular tilings they are based on are different. Three examples of $[p, q]$ symmetry groups are shown in Figure 2.3.

The fundamental region of $[p, q]$ is a right triangle with interior angles $\pi/p$ and $\pi/q$. The entire group can be said to be *generated* by reflections in the sides of this triangle, in the sense that every symmetry in the group is the product of a finite number of such reflections.

### *2.3 Tilings*

In their indispensible book *Tilings and Patterns* [68], Grünbaum and Shephard develop an extensive theory of tilings of the Euclidean plane. They begin from first principles with a nearly universally inclusive definition of tilings, one that permits so many pathological cases that the resulting objects cannot be meaningfully studied. They then proceed to layer restrictions upon the basic definition, creating ever smaller families of tilings that yield to more and more precise analysis and classification. Although the material persented here is largely drawn from *Tilings and Patterns*, it should not be considered to apply only in the Euclidean plane. Most of the basic facts about tilings apply equally well in non-Euclidean geometry.

For the purposes of creating the kinds of ornaments we will encounter in this work, we can jump in fairly late in the process and accept the following definition, corresponding to their notion of

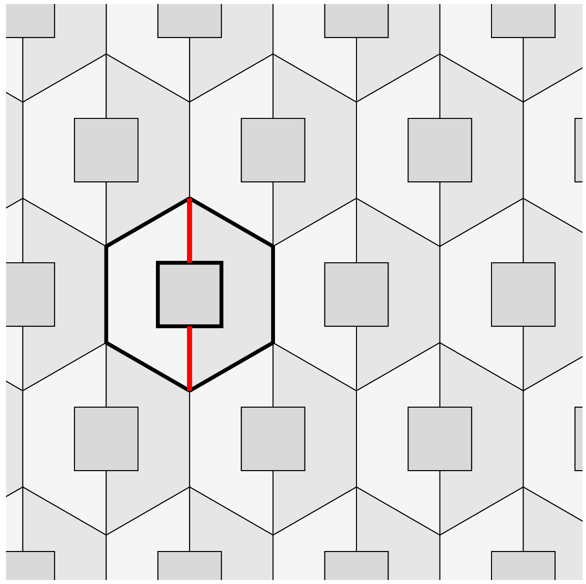


**Figure 2.4** A tiling for which some tiles intersect in multiple disjoint curves. The intersection between the two tiles outlined in bold is shown in red; it consists of two disjoint line segments.

"normal tiling" [68, Section 3.2].

**Definition (Tiling)** *A tiling is a countable collection $\mathcal{T}$ of tiles $\{T_1, T_2, \ldots\}$, such that:*

1. *Every tile is a closed topological disk.*
2. *Every point in the plane is contained in at least one tile.*
3. *The intersection of every two tiles is empty, a point, or a simple closed curve.*
4. *The tiles are uniformly bounded; that is, there exist $u, U > 0$ such that every tile contains a closed ball of radius $u$ and is contained in a closed ball of radius $U$.*

The most natural property associated with tilings, that they cover the plane with no gaps and no overlaps, is handled by conditions 2 and 3. Conditions 1 and 4 ensure that the tiles are reasonably well behaved entities that do not have exotic topological properties or become pathological at infinity.

Observe that condition 3 does more than prevent tiles from overlapping. It also prevents tilings like the one shown in Figure 2.4 from arising, where the boundary between two tiles is disconnected. When two tiles intersect in a curve, we may then refer to this well-defined curve as a *tiling edge*.

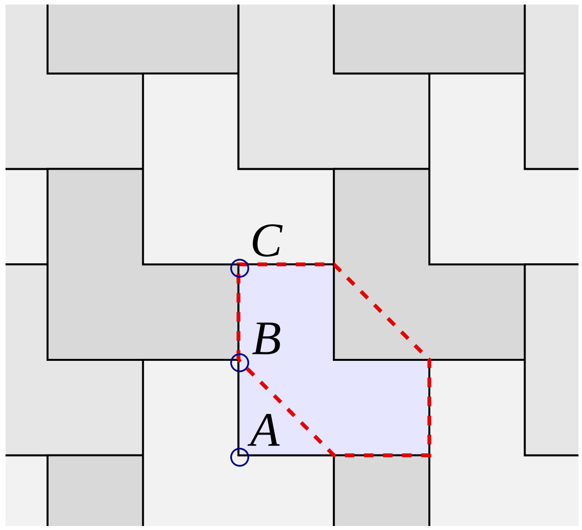



**Figure 2.5** The features of a tiling with polygonal tiles. For the tile highlighted in blue, $A$ is a shape vertex but not a tiling vertex, $B$ is a tiling vertex but not a shape vertex, and $C$ is both a tiling vertex and a shape vertex. The tiling polygon is shown as a dashed red outline.

Every tiling edge begins and ends at a *tiling vertex*, a place where three or more tiles meet. The tiling vertices are topologically important points in a tiling, as they determine the overall connectivity of the tiles and adjacencies between them. We will sometimes use the term "tiling polygon" to refer to the polygon joining the tiling vertices that lie on a single tile.

When the tiles in a tiling are polygons, there can be some confusion between the tiling vertices and edges as described above and the vertices and edges of individual polygons. To avoid confusion, we refer to the latter features when necessary as *shape vertices* and *shape edges*. Shape vertices and edges are properties of tiles in isolation; tiling vertices and edges are properties of the assembled tiling. Although the features of the tiling occupy the same positions as the features of the tiles, they may break down differently, as shown in Figure 2.5. When the two sets of features coincide (that is, when the tiling vertices are precisely the shape vertices), the tiling is called *edge-to-edge*.

In many of the tilings we see every day on walls and streets, the tiles all have the same shape. If every tile in a tiling is congruent to some shape $T$, we say that the tiling is *monohedral*, and that $T$ is the *prototile* of the tiling. More generally, a $k$-*hedral* tiling is one in which every tile is congruent to one of $k$ different prototiles. We also use the terms *dihedral* and *trihedral* for the cases $k = 2$ and $k = 3$, respectively. Note that a tiling need not be $k$-hedral for any finite $k$.

Just as with any other object in the plane, a tiling can be symmetric. Symmetries are a property of figures in the plane; the figure of a tiling can be taken as the set of points that lie on the boundary of any tile. Alternatively, we can alter the definition of symmetry slightly, saying that a symmetry of a tiling is a rigid motion that maps every tile onto some other tile.

### *2.3.1 Regular and uniform tilings*

A *regular tiling* is an edge-to-edge tiling of the plane by congruent regular polygons. In the Euclidean plane, an easy calculation shows that the only regular tilings are the familiar ones by squares, equilateral triangles, and regular hexagons.

We may then ask about edge-to-edge tilings of the plane using two or more different regular polygons as prototiles. In general, we can say very little about an arbitrary tiling of this type. To get more predictable behaviour, we further require that the vertices be indistinguishable, in the sense that any vertex can be mapped onto any other via a rigid motion that is also a symmetry of the tiling as a whole.[3] Given such a restriction, we can describe the tiling using a *vertex symbol*. A vertex symbol is a sequence $p_1.p_2 \ldots p_n$ that enumerates, in order, the regular polygons encountered around every vertex in the tiling. The tilings that can be described using vertex symbols in this way are called *uniform tilings*.

In the Euclidean plane, we can enumerate all vertex symbols that correspond to legal uniform tilings of the plane [68, Section 2.1]. The result is a set of eleven distinct tilings, sometimes also known as the *Archimedean tilings*. We name these tilings by placing their vertex symbols in parentheses. Among the uniform tilings the regular tilings are the ones whose vertex symbols are of the form $p^q$: $(4^4)$ for squares, $(3^6)$ for equilateral triangles, and $(6^3)$ for regular hexagons. (In a vertex symbol, we abbreviate blocks of repeated values using exponentiation.) Figure 2.6 shows the eleven Archimedean tilings.

There are vertex symbols that clearly cannot correspond to tilings of the Euclidean plane because

---

[3] Such tilings are usually called *isogonal*. In an isogonal tiling, the vertices form one transitivity class with respect to the tiling's symmetries. Isohedrality, a definition based on transitivity of whole tiles, will be encountered in Section 2.4.

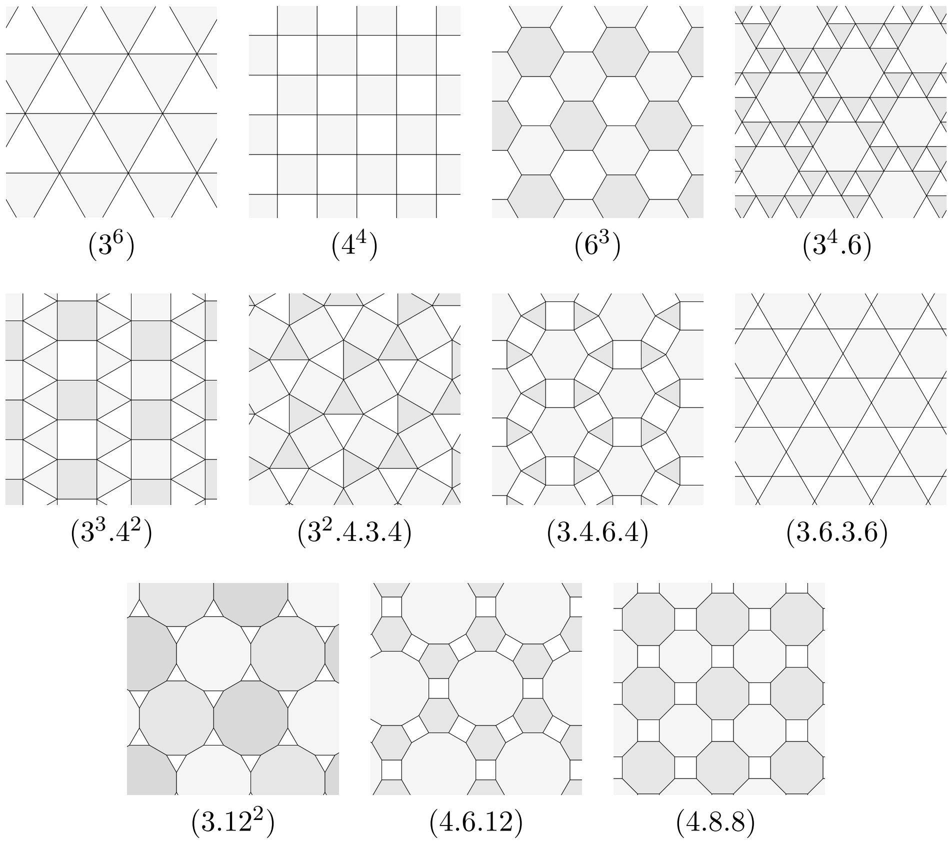



**Figure 2.6** The eleven uniform Euclidean tilings, also known as Archimedean tilings. The tiling $(3^4.6)$ occurs in two mirror-image forms.

the interior angles of the regular polygons around a vertex do not sum to 360 degrees. For example, five equilateral triangles leave a gap when arranged around a point. Yet there is a familiar shape with five equilateral triangles around every point: the icosahedron. Indeed, if we “inflate” the icosahedron so that all its edges become arcs of great circles on a unit sphere (line segments in spherical geometry), the result is a regular tiling of the sphere that can legitimately be called $(3^5)$.

In general, a given vertex symbol $p_1.p_2 \ldots p_n$ may describe a Euclidean, spherical, or hyperbolic tiling (or no tiling at all). If there is a tiling associated with the symbol, the tiling will be Euclidean, spherical, or hyperbolic when $\sum_{i=1}^{n} \frac{1}{p_i}$ is respectively equal to, greater than, or less than $\frac{1}{2}$. In particular, for any $p \geq 3$ and $q \geq 3$ there is a regular tiling $(p^q)$, consisting of regular $p$-gons meeting $q$ around every vertex. We have already seen the three Euclidean cases. The five spherical

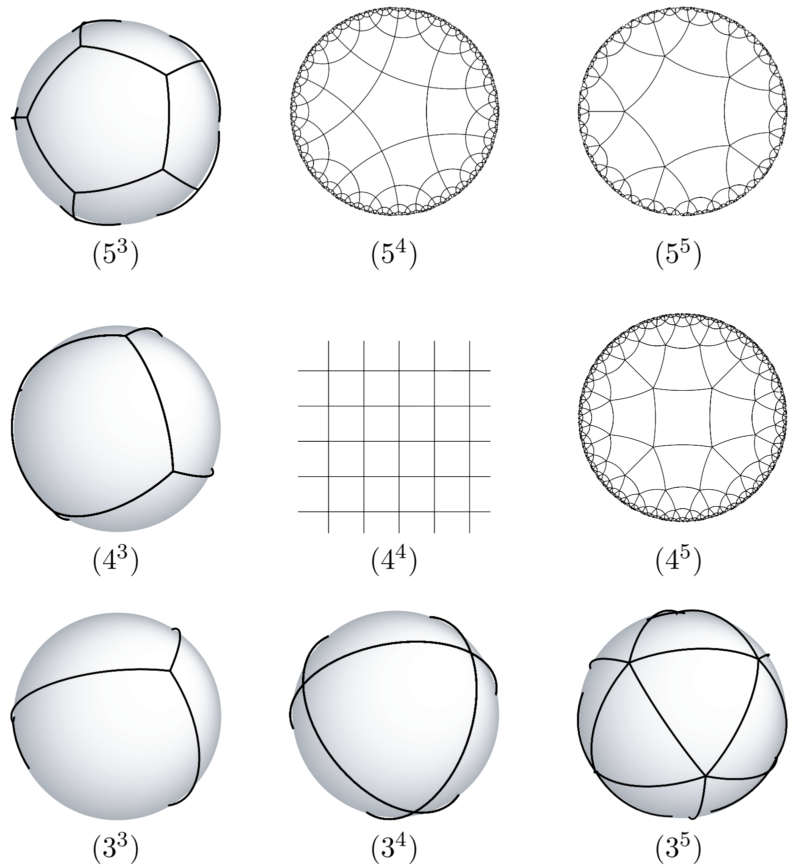



**Figure 2.7** Some Euclidean and non-Euclidean regular tilings.

cases correspond to the Platonic solids. The remaining cases are all hyperbolic. Some examples of regular tilings are shown in Figure 2.7. As discussed in Section 2.2.2, each regular tiling $(p^q)$ has symmetry group $[p, q]$, generated by reflections in a right triangle with interior angles $\pi/p$ and $\pi/q$.

There is no simple mathematical formula for deciding whether an arbitrary vertex symbol can be realized as a uniform tiling. For the most part, one simply tries to build the tiling, either discovering that all roads lead to contradictions, or finding a pattern that allows the tiling to be continued forever. As in the Euclidean case, the uniform tilings of the sphere are completely enumerated[4] – they

[4]The uniform tilings of the sphere should not be confused with the so-called *uniform polyhedra* [77]. A uniform polyhedron is allowed to have non-convex faces and may pass through itself. These polyhedra follow from a natural extension of the notion of uniform tiling in which the vertex symbol is allowed to have rational entries. There are

correspond to the five Platonic and thirteen Archimedean solids [136].

The existence of infinitely many regular tilings of the hyperbolic plane implies that there are infinitely many uniform tilings. Enumerating all possible uniform hyperbolic tilings remains an active area of inquiry. One approach is to define functions that transform a regular tiling into a related uniform tiling. Many of these functions are hyperbolic equivalents of spherical versions used by Hart to create novel polyhedra [75] (these functions implement what Hart calls "Conway notation"). Mitchell uses a similar idea to prove the existence of many parameterized families of hyperbolic uniform tilings [112].

Throughout the foregoing discussion, we have avoided the question of whether a given vertex symbol uniquely determines a uniform tiling. As Grünbaum and Shephard point out, there is no reason why this should necessarily be true, but it happens that in the Euclidean plane the uniform tilings are "just barely" unique [68, Page 64]. The same could be said for the sphere. In the hyperbolic plane, there is no guarantee of uniqueness; combinatorially distinct uniform tilings exist with the same vertex symbol [107]. The lack of uniqueness further complicates the enumeration of uniform hyperbolic tilings.

### *2.3.2 Laves tilings*

Every uniform tiling has a well-defined geometric dual, obtained by replacing every $n$-sided face by an $n$-valent vertex and vice versa. These dual tilings are monohedral and have the property that every tiling vertex is regular: the edges leaving the vertex are evenly spaced around it. In the Euclidean plane, these duals are called the *Laves tilings*, and they are given names analogous to their Archimedean duals. They are depicted in Figure 2.8. The Laves tilings will prove useful in Section 4.5, where they serve as a set of "defaults" from which to develop more complex tilings.

### *2.3.3 Periodic and aperiodic tilings*

A *periodic tiling* is a tiling of the Euclidean plane with periodic symmetry. That is, there exist two linearly independent directions of translational symmetry. In addition to a fundamental region, every periodic tiling has a *translational unit*, which is a fundamental region of the translational subgroup

---

seventy-five uniform polyhedra.

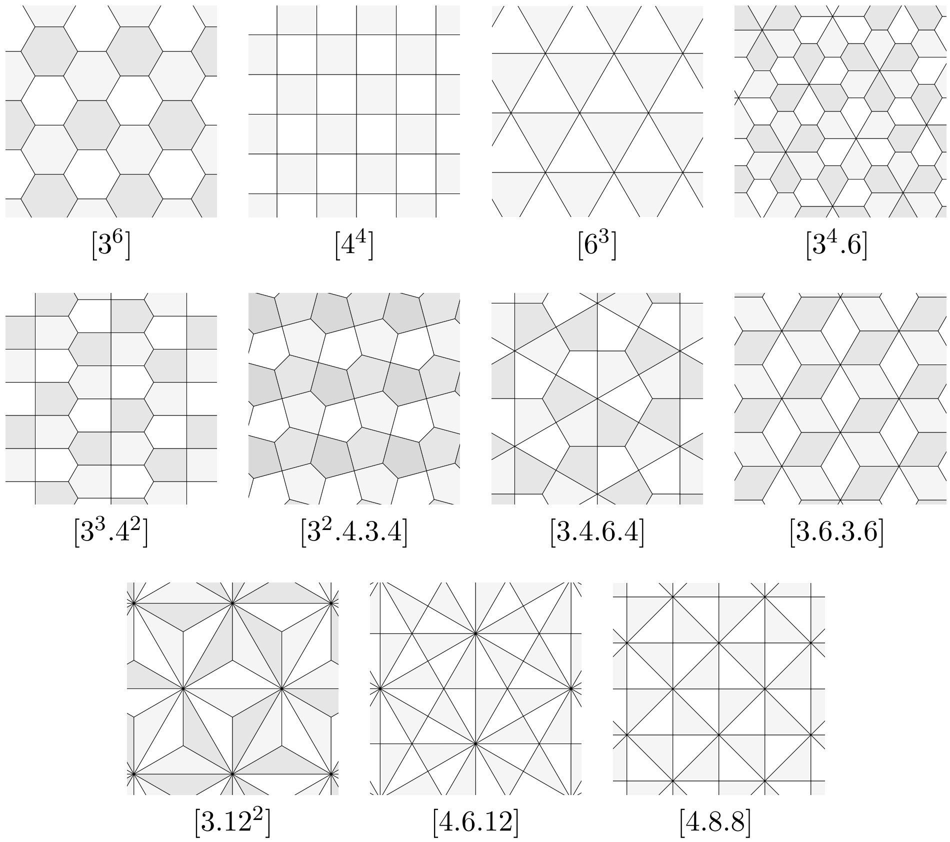
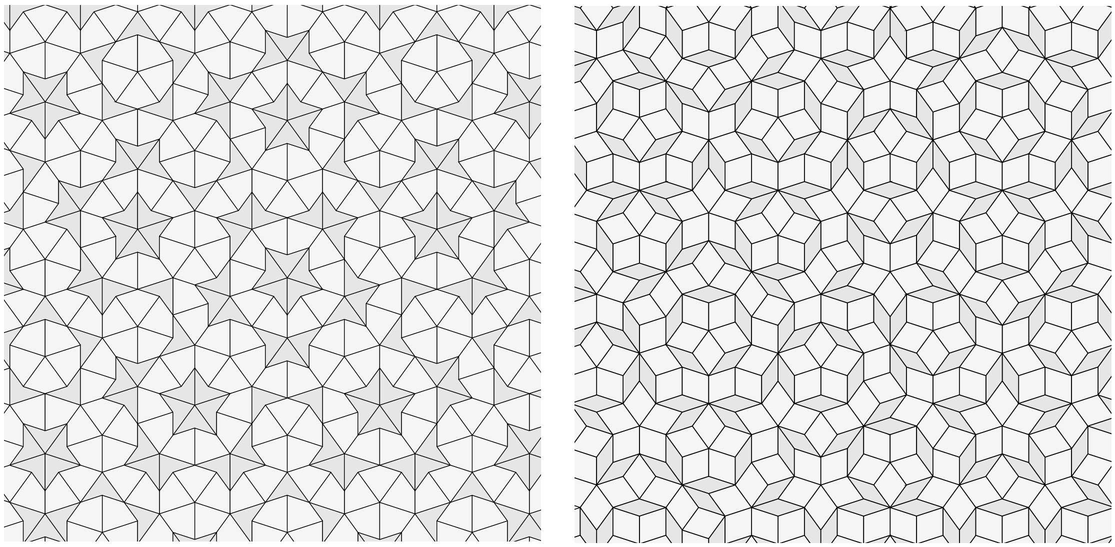



**Figure 2.8** The eleven Laves tilings. Each one is the dual of a corresponding Archimedean tiling. The tiling $[3^4.6]$ occurs in two mirror-image forms.

of the tiling's symmetry group. The translational unit can always be chosen to be a parallelogram with sides equal to two of the tiling's shortest non-parallel translational vectors.

The behaviour of periodic tilings is much more predictable than that of tilings in general because this behaviour only varies over a single translational unit. In a periodic tiling, a translational unit can be assembled from a finite number of tiles, guaranteeing that there are only finitely many different tile shapes. Each prototile can only occur in finitely many orientations and reflected orientations. We refer to these orientations collectively as the prototile's *aspects*, and distinguish the *direct aspects* from the *reflected aspects* when necessary.

At the opposite end of the spectrum, *aperiodic tilings* have received a great deal of attention over the past few decades, both from tiling theorists and lay audiences. Especially popular are Penrose's



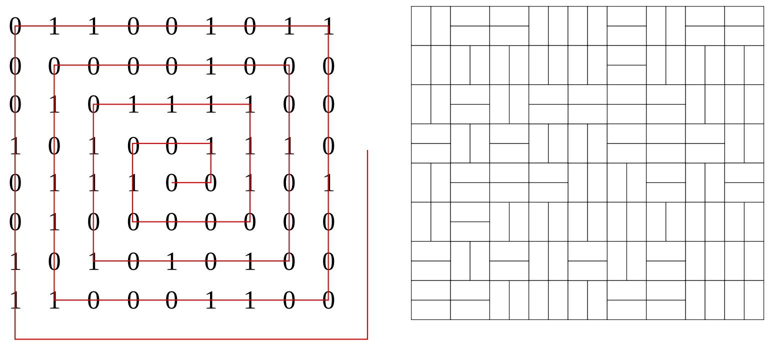
(a) (b)

**Figure 2.9** The two famous aperiodic tilings of Penrose. The "kite and dart" tiling is shown in (a) and thin and thick rhombs in (b).

aperiodic tilings, which have provided a steady stream of results and inspiration in physics [74, Section 9.8], algebraic geometry [23], and the popular mathematical press [119, Chapter 7]. Of the aperiodic tilings devised by Penrose, the most well known are the ones shown in Figure 2.9, made up either of "kites" and "darts" or thin and thick rhombs.

Like the Mandelbrot set, the Penrose tilings are ambassadors of the geometric aesthetic, bringing their message of mathematical beauty to a general audience. A quick survey of the Geometry Junkyard's web page on Penrose tilings [46] shows dozens of non-technical ways that they have been applied. They have been fashioned into puzzles, games, and fridge magnets; used to cover walls, floors, and windows; even sewn into lace doilies. Glassner has discussed the creation of graphical ornament based on Penrose tilings [54, 55].

A tiling that is not periodic is called *nonperiodic*. A frequent but incorrect assumption is that the definitions of aperiodic and nonperiodic coincide. In fact, the aperiodic tilings are a very special subset of the nonperiodic tilings. It is worthwhile to clarify the distinction between the two, showing why the aperiodic tilings are an active and exciting area of research.



(a) (b)

**Figure 2.10** A contrived example of how even a very simple shape may yield aperiodic tilings. A spiral path is used to place digits from the binary expansion of $\pi$, as given by Wolfram [138, Page 137]. Each digit is then used to place a pair of bricks, oriented vertically to represent a 0 and horizontally to represent a 1. The resulting tiling, when extended to the whole plane, is clearly aperiodic, even though the brick prototile could easily be used to construct periodic tilings. There exist uncountably many aperiodic tilings based on this prototile.

Lack of periodicity is not in itself a very surprising property. One might argue that nonperiodicity is the common case in the universe of tilings, and that the periodic tilings represent small pockets of order in a sea of chaos. This situation is reminiscent of the disposition of the rational numbers, peppered among the much more common irrationals. This analogy can be interpreted quite literally: a number is rational precisely when its digit sequence is periodic. To move from this chaos to a definition of aperiodicity, we must must narrow our point of view significantly.

It is easy to construct tilings where every tile shape is unique – consider, for example, the Voronoi diagram induced by an infinite integer lattice whose points have been jittered randomly. Clearly there can be no hope of periodicity in such a case, but this fact seems unimpressive. In developing a definition of aperiodicity, we therefore consider only those nonperiodic tilings with a finite number of prototiles, *i.e.*, those that are $k$-hedral for some $k$.

Under this restricted definition, we can still construct very simple nonperiodic tilings. Even a 2-by-1 brick yields an infinite variety, as demonstrated in Figure 2.10. These tilings seem contrived, however, because the same brick can easily be made to tile periodically. Nonperiodic tilings become

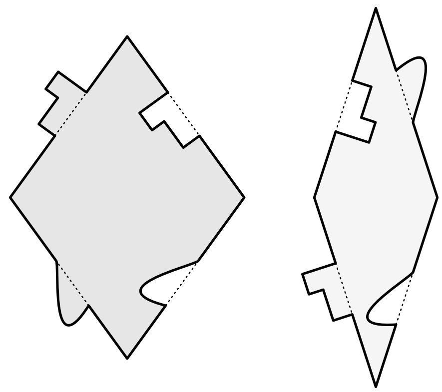


**Figure 2.11** Sample matching conditions on the rhombs of Penrose's aperiodic tile set $P3$. The unmodified rhombs (indicated by dotted lines) can form many periodic tilings. The puzzle-piece deformations on the tile edges guarantee that any tiling formed from these new shapes will be aperiodic.

truly interesting when we take into account all possible alternative tilings that can be constructed from the same set of prototiles. We call a set of prototiles an *aperiodic tile set* when every tiling admitted by that set is nonperiodic. (We also ask that the set admit at least one tiling!) We can then define an *aperiodic tiling* as a tiling whose prototiles are an aperiodic tile set. An aperiodic tiling is one that is "essentially nonperiodic," in the sense that no rearrangement of its tiles will achieve periodicity. When used to refer to a particular tiling, aperiocity is therefore a far reaching concept — it encompasses all possible tilings that can be formed from the same prototiles.

The Penrose tilings highlight the special behaviour of aperiodic tilings. Consider the tiling by Penrose rhombs shown in Figure 2.9(b). By themselves, the two rhombs do not form an aperiodic tile set; they can be arranged into both periodic and nonperiodic tilings. What makes the rhombs aperiodic are additional "matching conditions" that are imposed on them, limiting the ways that two tiles may be adjacent to one another. These matching conditions can be expressed in a number of ways. One possibility is to modify the shapes of the tiles by adding protrusions to the rhomb edges, so that the tiles must snap together like pieces in a jigsaw puzzle [68, Section 10.3] (we will exploit these geometric matching conditions in Section 4.6). A set of protusions that express the matching conditions is shown in Figure 2.11. It is these two modified shapes that form an aperiodic tile set, known as the Penrose tile set $P3$ (the modified kite and dart are known as $P2$). Many sets of prototiles must be endowed with similar matching conditions to enforce aperiodicity. The matching conditions are typically not shown when the tilings are rendered, perhaps leading to the confusion between nonperiodicity and aperiodicity.

As it turns out, Penrose tilings have even more structure than that determined by the aperiodicity of their prototiles. A beautiful argument shows that even though there are infinitely many different combinatorially inequivalent Penrose tilings from a given set of prototiles, they are all indistinguishable on any bounded region of the plane. Given some Penrose tiling, if we let $\mathcal{P}$ denote the set of tiles that lie inside some bounded region of the plane, then $\mathcal{P}$ will occur infinitely often in every Penrose tiling by those same prototiles. Tilings like these seem to live right on the edge of periodicity, and receive the special designation *quasiperiodic*.

For many years, research in aperiodic tilings has sought to create ever smaller aperiodic tile sets. The first aperiodic sets contained thousands of prototiles [68, Chapter 11]. Penrose managed to bring the minimum number of prototiles needed down to two. Other small aperiodic tile sets have been discovered by Ammann [68, Section 10.4] and Goodman-Strauss [61]. However, nobody has been able to improve upon this result or to show that at least two tiles shapes are required. The question of whether there exists an *aperiodic tile*, a single shape that tiles only aperiodically, remains open [28, Section C18], and is one of the most beautiful unsolved problems in geometry.

## *2.4 Transitivity of tilings*

For two congruent tiles $T_1$ and $T_2$ in a tiling, there will be some rigid motion of the plane that carries one onto the other (there may in fact be several). A somewhat special case occurs when the rigid motion is also a symmetry of the tiling. In this case, when $T_1$ and $T_2$ are brought into correspondence, the rest of the tiling will map onto itself as well. We then say that the two tiles are *transitively equivalent*.

Transitive equivalence is an equivalence relation that partitions the tiles into *transitivity classes*. When a tiling has only one transitivity class, we call the tiling *isohedral*. More generally, a $k$-isohedral tiling has $k$ transitivity classes. An isohedral tiling is one in which a single prototile can cover the entire plane through repeated application of rigid motions from the tiling's symmetry group. In an isohedral tiling, there is effectively no way to tell any tile from any other.

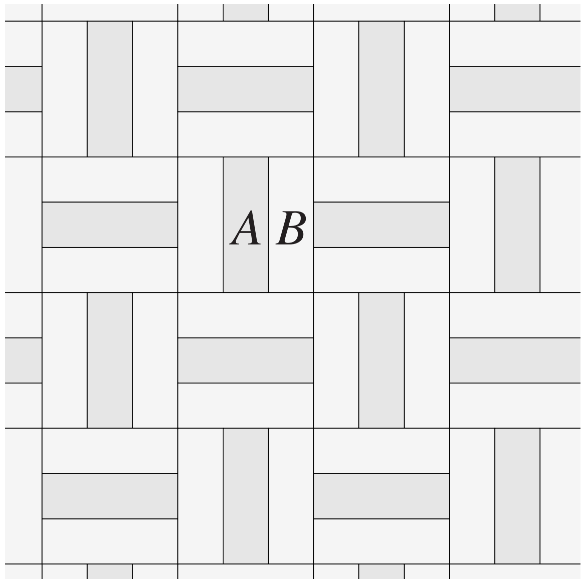



**Figure 2.12** An example of a monohedral tiling that is not isohedral. The two tiles labeled $A$ and $B$ cannot be in the same transitivity class, which can be seen by the different ways each is surrounded by its neighbours. This tiling is 2-isohedral.

Two tiles in the same transitivity class must obviously be congruent, but the converse need not be true. Figure 2.12 shows a monohedral tiling with two transitivity classes. The two classes of tiles can be distinguished by the arrangement of a tile's neighbours around it.

As with the definition of aperiodicity, while the tiling given above is 2-isohedral, it seems that way only in a weak sense, because the same shape also tiles the plane isohedrally. We must ask, therefore, whether there exists an *anisohedral tile*, a single shape that tiles the plane, but never isohedrally. In 1900, Hilbert seemed to take it for granted that no such shape can exist [68, Section 9.6]. In 1935, however, Heesch demonstrated an anisohedral prototile [79], reproduced in Figure 2.13.

In Heesch's example, a single prototile generates tilings with two transitivity classes. We therefore specify that the given tile is *2-anisohedral*, and generalize the definition, calling a prototile $k$-*anisohedral* if every tiling it admits has at least $k$ transitivity classes. Grünbaum and Shephard exhibit a collection 2- and 3-anisohedral pentagons, and ask for what values of $k$ there exist $k$-anisohedral tiles [68, Section 9.3]. In 1993, Berglund found a 4-anisohedral prototile, and called for examples with $k \geq 5$ [10]. Since 1996, using computer searches over polyominoes, polyia-

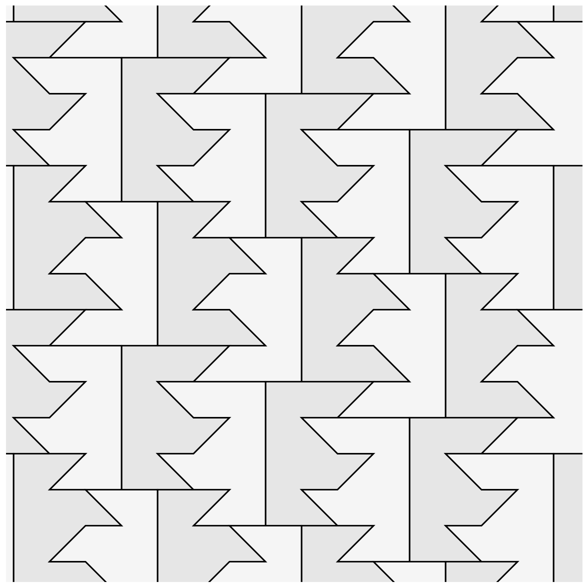


**Figure 2.13** Heesch's anisohedral prototile. No tiling that can be assembled from this shape will be isohedral.

monds, and polyhexes,[5] Joseph Myers has demonstrated $k$-anisohedral tilings for all $k \leq 9$ [115]. There seems to be no reason to assume an upper bound on possible values of $k$, though one must imagine the search will become more difficult each time $k$ increases. The search for $k$-anisohedral prototiles is related to the search for an aperiodic prototile, in the sense that an aperiodic prototile is $\infty$-anisohedral.

### *2.4.1 Isohedral tilings*

By definition, an isohedral tiling is bound by a set of geometric constraints: congruences between tiles must be symmetries of the tiling. Grünbaum and Shephard show that those geometric constraints can be equated with a set of *combinatoric* constraints expressing the adjacency relationship a tile maintains across its edges with its neighbours. They prove that the constraints yield a divi-

[5] A *polyomino* is the union of a finite set of connected squares from the regular tiling by squares. *Polyiamonds* and *polyhexes* can be defined analogously from the regular tilings by triangles and hexagons. A general introduction to the field is given by Golomb [59], who coined the term; Grünbaum and Shephard discuss polyominoes as tilings of the plane [68, Section 9.4].

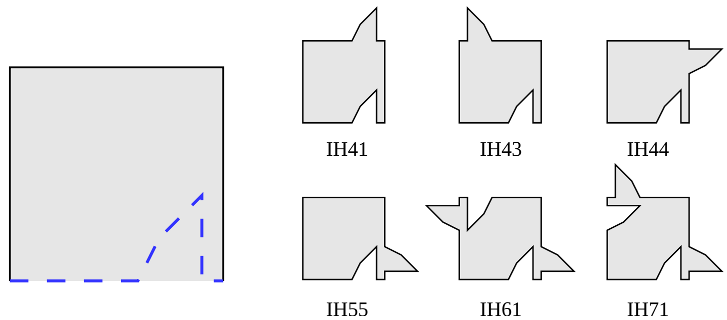



**Figure 2.14** An isohedral tiling type imposes a set of adjacency constraints on the tiling edges of a tile. When the bottom edge of the square deforms into the dashed line, the other edges must respond in some way to allow the new shape to tile. The six resulting prototiles tiles here are from six different isohedral types, and show six of the possible responses to the deformation.

sion of the isohedral tilings into precisely 93 distinct *types* or *families*,[6] referred to individually as IH1, ..., IH93 and collectively as IH [68, Section 6.2]. Each family encodes information about how a tile's shape is constrained by the adjacencies it is forced to maintain with its neighbours. In 12 of these types, the adjacency relationship can only be realized by placing *markings* on tiles that indicate their orientations. We will primarily be concerned with the other 81 types, where the combinatoric structure of the tiling can be expressed geometrically through deformations to the tiling edges. A change to a tiling edge is counterbalanced by deformations in other edges; which edges respond and in what way is dependent on the tiling type, as shown in Figure 2.14. In what follows, we review the classification and notation used with the isohedral tilings.

The *combinatorial structure* of an isohedral tiling $\mathcal{T}$ is an infinite graph whose vertices are the tiling vertices of $\mathcal{T}$, and where two vertices are connected by an edge if the two corresponding tiling vertices are connected by a tiling edge. Two isohedral tilings can then be said to be *combinatorially equivalent* when their combinatorial structures are isomorphic. Combinatorial equivalence partitions the isohedral tilings into eleven classes, referred to as *combinatorial types*, or more commonly as *topological types*.[7] Each topological type has one of the eleven Laves tilings as a distinguished

[6]In tiling theory, seemingly arbitrary numbers like 93 are not uncommon; enumerations of families of tilings tend to have sets of constraints that collapse certain cases and fracture others.

[7]The use of the term "topological type" would seem to suggest that the two tilings are topologically, and not combina-

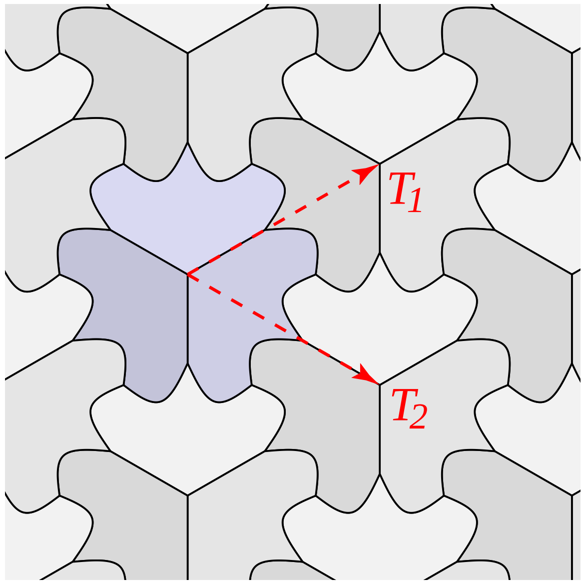



**Figure 2.15** An example of an isohedral tiling of type IH16. A single translational unit of the tiling is shown through the two translation vectors $T_1$ and $T_2$ and the three coloured aspects.

representative, and we name the type using the vertex symbol of the corresponding Laves tiling. For example, Figure 2.15 shows an isohedral tiling of type IH16. We can see that every tile has six tiling vertices, all of valence three, meaning that IH16 is of topological type $3^6$.

Every isohedral tiling is both monohedral and periodic, meaning that its behaviour over the entire plane can be summarized by specifying the aspects of the single prototile that make up a translational unit, and two linearly-independent translation vectors that replicate that unit over the plane. IH16 has three aspects, shown in varying shades of blue in Figure 2.15. These three tiles comprise one possible translational unit with translation vectors $T_1$ and $T_2$.

The adjacency constraints between the tiling edges of a tile are summarized by an *incidence symbol*. Given a rendering of an isohedral tiling, the incidence symbol can be derived in a straightforward way.

Figure 2.16 shows five steps in the derivation of an incidence symbol for our sample tiling. To

torially equivalent. In fact, for normal tilings the two forms of equivalence are identical [68, Section 4.1] and so both terms are valid.

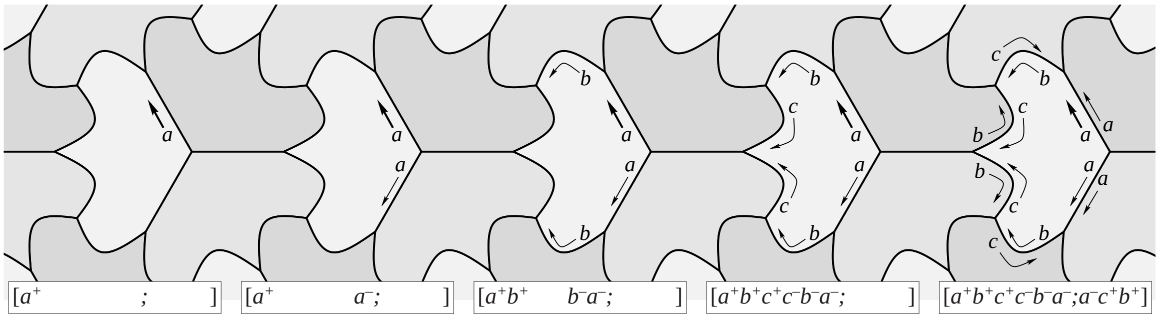



**Figure 2.16** Five steps in the derivation of an isohedral tiling's incidence symbol.

obtain the first part of the incidence symbol, we pick an arbitrary tiling edge as a starting point, assign that edge a single-letter name, and draw an arrow pointing counterclockwise around the tile (step 1). Then, we copy the edge's label to all other edges of the tile related to it through a symmetry of the tiling (step 2). Should the edge get mapped to itself with a reversal of direction, it becomes undirected and is given a double-headed arrow. We then proceed counterclockwise around the tile to the next unlabeled edge (if there is one) and repeat the process (step 3). The first half of the symbol is obtained by reading off the assigned edge names (step 4). A directed edge is superscripted with a sign indicating the agreement of its arrow with the traversal direction. Here, a plus sign is used for a counterclockwise arrow and a minus sign for a clockwise arrow.

The second half of an incidence symbol records how, for each different label, a tiling edge with that label is related to the corresponding edge of the tile adjacent to it. To derive this part of the symbol, we copy the labeling of the tile to its neighbours (step 5). Then, for each unique edge letter assigned in the first step, we write down the edge letter adjacent to it in the tiling. If the original edge was directed, we also write down a plus or minus sign, depending on whether edge direction is respectively preserved or reversed across the edge. A plus sign is used if the arrows on the two sides of a tiling edge are pointing in opposite directions, and a minus sign is used otherwise. For the running example, the incidence symbol turns out to be $[a^+b^+c^+c^-b^-a^-; a^-c^+b^+]$. Note that the incidence symbol is not unique; edges can be renamed and a different starting point can be chosen. But it can easily be checked whether two incidence symbols refer to the same isohedral type.

Every isohedral type is fully described in terms of a topological type and an incidence symbol.

Enumerating all possible topological types and incidence symbols and eliminating the ones that do not result in valid tilings or that are trivial renamings of other symbols leads to the classification given by Grünbaum and Shephard.

### 2.4.2 *Beyond isohedral tilings*

Since the work of Grünbaum and Shephard on the classification of isohedral tilings of the Euclidean plane, other tiling theorists have gone on to search for generalizations to related tilings. In particular, a group led by Dress, Delgado-Friedrichs, and Huson pioneered the use of *Delaney symbols* in the study of what they call *combinatorial tiling theory* [33, 34, 86]. A Delaney symbol completely summarizes the combinatorial structure of a $k$-isohedral tiling of the Euclidean plane, the hyperbolic plane, or the sphere. They can also be generalized to tilings in spaces of dimension three and higher. Delaney symbols form the basis for an efficient software implementation, and Delgado-Friedrichs and Huson have created **2dtiler**, a powerful tool for exploring, rendering, and editing tilings from their combinatorial descriptions.

Combinatorial tiling theory does not play a direct role in the present work (although it is used in a classification by Dress that motivates the technique of Section 4.6.2). Nevertheless, Delaney symbols have helped to advance tiling theory beyond the material of *Tilings and Patterns*, and it seems likely that adopting them as a standard description of all-over tilings could lead to principled $k$-isohedral and non-Euclidean generalizations of the algorithms and data structures presented in Chapter 4.

## ***2.5 Coloured tilings***

Up to now, we have ignored the possibility of colouring tiles in a tiling. When analyzing the symmetries of a tiling, we have treated colour as superficial, to be disregarded when deciding whether two tiles are "the same." It is also possible to take colour into account, adding a layer of richness and complexity to a tiling. The colouring can have a great deal of structure, particularly when it acts compatibly with the symmetries of the tiling. Coxeter gives a group-theoretic presentation of colouring, using Escher's tilings as motivating examples [24]. Grünbaum and Shephard provide an extensive account of the relationship between colouring and tilings [68, Chapter 8]. We restate two important definitions here.

A $k$-*colouring* of a tiling is a function from tiles to the set $\{1, \ldots, k\}$ that assigns an abstract "colour" to each tile.[8] That colouring is a *perfect colouring* if every symmetry of the tiling acts as a permutation of the colours. Symbolically, let $\sigma$ be a symmetry of a tiling $\mathcal{T}$ with colouring $c : \mathcal{T} \to \{1, \ldots, k\}$. The rigid motion $\sigma$ maps every tile $T \in \mathcal{T}$ to some tile $\sigma(T)$. Then the colouring $c$ is a perfect colouring if for any symmetry $\sigma$, there exists a permutation $\rho$ of the set $\{1, \ldots, k\}$ such that $c(\sigma(T)) = \rho(c(T))$ for all $T \in \mathcal{T}$.

Escher studied colourings of tilings in depth while preparing his notebook drawings. He paid great attention to the question of colouring, expressing as a clear objective that adjacent tiles should have contrasting colours to better distinguish them from each other [50]. In general, he aimed to achieve this contrast with a minimal number of colours. Yet his intuition seems to have guided him to the perfect colourings, in some cases choosing a perfect colouring with more colours over a non-perfect one with fewer. A clear example is symmetry drawing 20 [124, Page 131], where a tiling coloured perfectly by four colours is accompanied by a note mentioning that three would have sufficed to distinguish adjacent tiles. Shephard points out that for this tiling, no perfect colouring is possible with only three colours [126]. Escher intuited that a fourth colour allowed for a more regular colouring. Four colours suffice to perfectly colour any isohedral tiling in such a way that adjacent tiles have contrasting colours.

Escher's understanding of the compatibility between a tiling's symmetries and its colouring predated the development of a formal theory of colour symmetry, and to some extent set that development in motion [123]. For while a small amount of mathematical work had been done on the subject previously, it was when the crystallography community became aware of Escher's tesselations that they understood how much of a theory there was to be had, and they were provided with a rich library of illustrations from which to build that theory.

---

[8] This definition of $k$-colouring should be distinguished from its use in problems of graph colouring and map colouring, where there is an additional restriction that adjacent vertices or regions have different colours (as in the four colour theorem). A $k$-colouring does not require adjacent tiles to have different colours, although we will discuss this additional restriction as well.