Symmetries of Venn Diagrams on the Sphere

by

Mark R. N. Weston
BSc, University of Victoria, 2000
MSc, University of Victoria, 2003

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy

in the Department of Computer Science

Symmetries of Venn Diagrams on the Sphere

by

Mark R. N. Weston
BSc, University of Victoria, 2001
MSc, University of Victoria, 2003

Supervisory Committee

Dr. Frank Ruskey, Supervisor (Department of Computer Science)

Dr. Wendy Myrvold, Member (Department of Computer Science)

Dr. Valerie King, Member (Department of Computer Science)

Dr. Gary MacGillivray, Outside Member (Department of Mathematics)

Supervisory Committee

Dr. Frank Ruskey, Supervisor (Department of Computer Science)

Dr. Wendy Myrvold, Member (Department of Computer Science)

Dr. Valerie King, Member (Department of Computer Science)

Dr. Gary MacGillivray, Outside Member (Department of Mathematics)

Abstract

A diagram on a surface is a collection of coloured simple closed curves which generally intersect only at points, and a Venn diagram, of n curves, has the additional property that there are exactly 2^n faces in the diagram, each corresponding to a unique intersection of the interiors of a subset of the curves. A diagram has plane symmetry if it can be constructed by rotating a single closed curve in the plane n times, each time by $2\pi/n$, and changing the colour of the curve for each rotation; equivalently, the diagram can be constructed from a region forming a "pie-slice" of the diagram and containing a section of each curve, and then copying and rotating this region n times, recolouring the sections of curves in the region appropriately. This and reflective symmetries are the only non-trivial ways a finite plane diagram can have some kind of symmetry.

In this thesis, we extend the notion of planar symmetries onto the sphere, by constructing and projecting diagrams onto the sphere and examining the much richer symmetry groups that result. Restricting our attention to Venn diagrams gives a combinatorial structure to the diagrams that we examine; we also derive several constructions of Venn diagrams with interesting symmetries on the sphere by modifying the landmark work of Griggs, Killian and Savage from 2004, which provided some important answers to questions about planar symmetric diagrams. This thesis also contains an examination of a class of diagrams that exhibit a rotary reflection symmetry (a rotation of the sphere followed by a reflection), in which we make some initial steps towards a general construction for n-Venn diagrams realizing a very rich symmetry group of size $2n$, for n prime or a power of two. We also provide a many-dimensional construction of very simple Venn diagrams which realize any subgroup of an important type of symmetry group that use only reflection symmetries. In summation, we exhibit and examine at least one Venn diagram realizing each of the 14 possible finite symmetry groups on the sphere. All of these investigations are coupled with a theoretical and practical framework for further investigation of symmetries of diagrams and discrete combinatorial objects on spheres and higher-dimensional surfaces.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I wish to acknowledge the patient guidance and support provided by my long-term supervisor Dr. Frank Ruskey, who has always been supportive and available in my various research endeavours, and without whom this research would never have been conducted.

I also wish to thank my collaborator Dr. Brett Stevens (Carleton), whose enthusiasm and dedication to these topics while visiting UVic in early 2007 inspired several of the chapters contained herein.

This research was supported in part by NSERC and by the University of Victoria's fellowship program.

Finally, I wish to thank my family and my wife, who have always cheerfully supported my scholastic pursuits.

Chapter 1

Introduction

The human figure has a bilateral symmetry—standing upright, most normally-developed people exhibit an almost-perfect left-to-right mirror symmetry. The reason for this is simple; on the surface of the earth there is no preference for moving in a particular direction, and we have evolved (modulo a few notable exceptions, such as left- and right-handedness) to be able to work, move and interact to either side of our bodies with equal ease. Our brains reflect this in our innate aesthetic and functional preference for symmetries of many types: so much analysis of architecture, painting, music, and other hard and soft arts revolves around the presence or absence of symmetry that it is difficult to argue that considerations of symmetry do not play a central component of our modern conception of beauty, harmony, balance, and perfection. And, of course, our modern culture of science, from the physicist's notions of space and time to the chemist's molecular models, is deeply imbued with the language of symmetry, grounded in solid mathematical foundations.

The concept of symmetry is fundamental to our study of many types of mathematical objects. The discipline that encompasses and formalizes this area is called group theory, and with its development by Galois and others in the 19th century the mathematical community could define and categorize different types of symmetry and so provide a solid theoretical foundation for the increasingly-important role of symmetry in other sciences.

> Numbers measure size, *groups measure symmetry.* [3, p. vii]

The story of groups has been told many times in popular, as well as scientific, literature—a good introduction, along with the dramatic stories of Galois and Abel, is found in [85]. Soon after their formalization, groups were being used by mathematicians such as Klein and Jordan to study geometric transformations. Klein, in 1872, proposed in his "Erlangen programme" that group theory was the way to formulate and understand geometric constructions, and the application of groups to geometry is still an active area of research, especially in hyperbolic geometry [8]. Many mathematicians in the 18th century, such as Gauss and Lagrange, produced important work that only in the 19th century and beyond was recognized as being essentially motivated by group-theoretic concepts [81]. Groups are found everywhere in mathematics and elsewhere, and they provide the theoretical foundation for our study of symmetry. The geometry and symmetries of figures in 2- and 3-dimensional space are often used as a context in which to present and elucidate group theoretic concepts, at levels from undergraduate to the advanced graduate [12, 99].

In the topic of information visualization, the role of diagrams of various types has been of prime importance. Logical diagrams as an aid to reasoning were originally developed as part of a lesson plan from the 18th-century mathematical genius Leonard Euler to a pupil [39], and, a century later, were developed and formalized by John Venn [125]. They form an important class of combinatorial objects that are now used in set theory and many applied areas, and are often taught at the grade-school level as a method of understanding simple logic. Of the few mathematicians able to bridge the divide between combinatorics and formal geometric and aesthetic considerations, Branko Grünbaum was a pioneer in formalizing the study of Venn diagrams [56], and he and other diagrammatical researchers began, in the last 50 years, to formally study them as mathematical objects worthy of consideration in their own right. Edwards' book [34] provides an introduction written for the non-scientific reader with a focus on the geometrical aspects of Venn diagrams and Edwards' own research. A more comprehensive historical survey will follow in subsequent chapters.

Applying the concept of symmetry to diagrams, especially Venn diagrams, has provided rich and fruitful areas of study resulting in the construction of fantastic new works of art; indeed, we know of at least one case of symmetric Venn diagrams being contextualized as mathematically-derived works of art and sold at an art show [21, 66]. Figure 1.1 shows an 11-curve symmetric Venn diagram with 2048 regions; each region is coloured according to its distance from the central region. This diagram exhibits a type of symmetry on the plane that we will further explore in this thesis. Moreover, the interest in symmetric diagrams lies both in their pleasing aesthetic qualities and their important combinatorial and group-theoretic properties; knowing that a diagram is symmetric imposes many constraints on the underlying group and combinatorial structure.

The interested reader can see many examples of (symmetric and asymmetric) Venn and Euler diagrams in an online dynamic survey [106] that presents many of the current research topics in Venn diagrams.

1.1 Overview

Research in symmetry of Venn diagrams has, with a few notable exceptions that we discuss in Chapter 3, been almost exclusively focussed on diagrams that have rotational symmetry in the plane. Moreover, the rotational symmetry studied has been of a very specific type: with a diagram composed of n curves, the symmetry allowed is one that maps each curve, in its entirety, to another, by rotating the entire diagram about a central point, contained in the central face.

In this thesis we wish to break out of the constricting milieu of rotational symme-

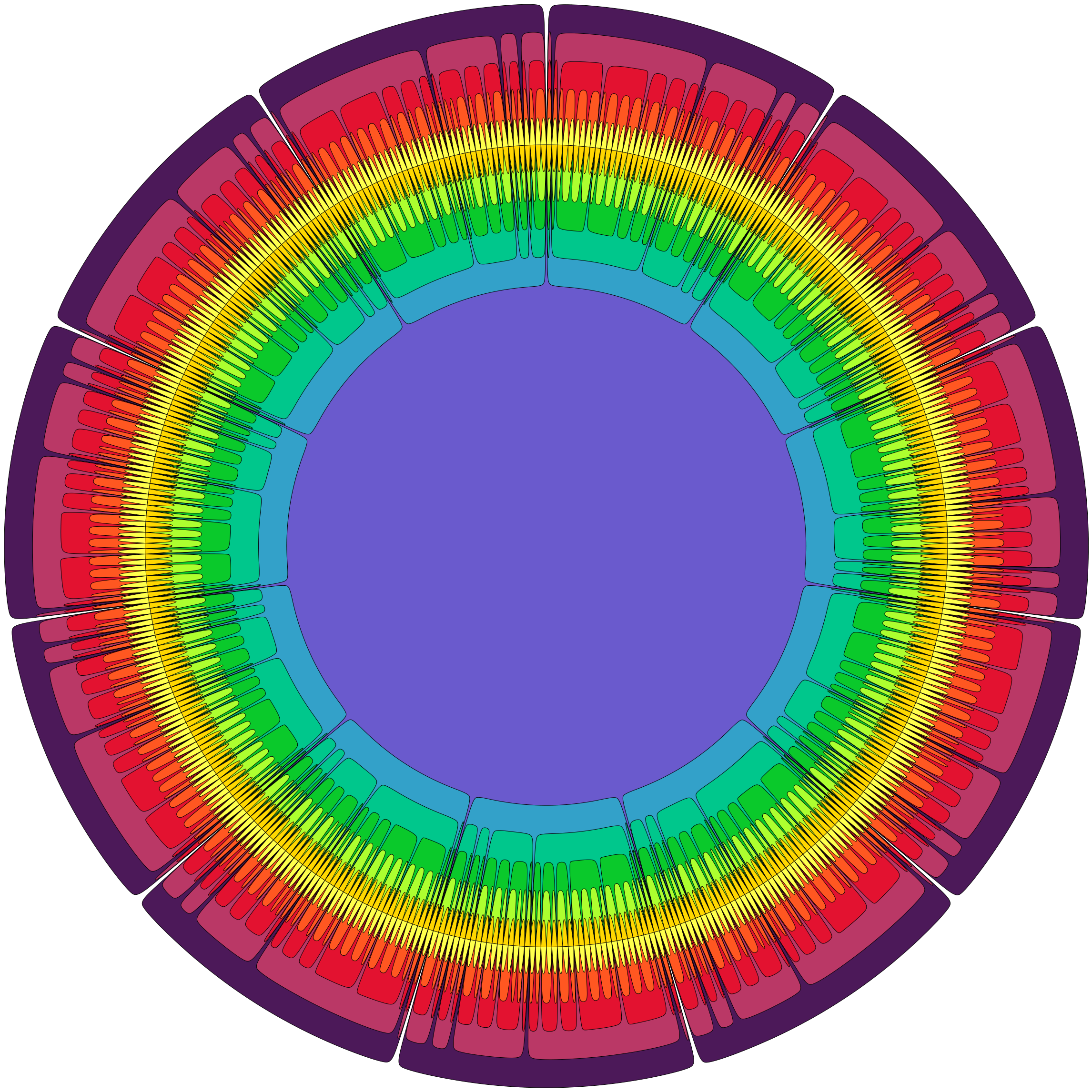


Figure 1.1: A beautiful diagram constructed by purely mathematical considerations of symmetry

tries on the plane, chiefly by examining symmetries of diagrams on the sphere. The main contributions of this work are several constructions for chain decompositions of posets and Venn diagrams that have non-trivial symmetry groups on the sphere, as well as a survey of many existing and new diagrams with symmetries on the sphere, and some results on higher-dimensional Venn diagrams; in total we exhibit diagrams for each of the 14 possible types of finite symmetry groups on the sphere.

Chapter 2 presents necessary definitions and introductory material for the rest of the work, including some more historical background, and Chapter 3 focusses on introductory material and previous work in symmetry in diagrams. Chapter 4 provides a framework for the representation and discussion of symmetric diagrams on the sphere, including colour symmetry as applied to diagrams. Chapter 5 introduces a method of producing symmetric chain decompositions of the boolean lattice with certain symmetries that gives two constructions for Venn diagrams on the sphere. Chapter 6 presents a construction for higher-dimensional Venn diagrams realizing any instance of a certain type of symmetry group on the higher-dimensional sphere. Chapter 7 contains some material on shift register sequences on boolean strings and their connections to diagrams on the sphere with a type of rotational symmetry, culminating in several 5-, 7-, and 8-curve diagrams exhibiting this rich symmetry group. Chapter 8 contains discussions of many other diagrams with different symmetry groups on the sphere, including discussion of a well-known construction for Venn diagrams by Edwards. Finally, in Chapter 9, Table 9.1 on page 235 presents a succinct encapsulation of all of the results of the thesis, by showing a compendium of all of the diagrams discussed in the thesis with symmetries on the sphere. The final chapter also presents some conclusions along with future directions of research.

Chapter 2

Background

In this chapter we shall cover the necessary definitions and basic prior results required for the rest of the work, as well as some historical information regarding the study of Venn diagrams. We follow the definitions of Grünbaum [56] and Ruskey and Weston [106], with a few specific deviations noted where they occur. Our source for topological definitions, and a reference for more detail on the foundations that we build on, is Henle [73].

2.1 Diagrams

The diagrams that we will be considering are (mostly) embedded on the the sphere or the plane. To begin to discuss curves accurately we must consider some notions of sets of points in these spaces. The next definitions follow standard practise.

Definition. The *plane* is the infinite set of points $\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$.

Three-dimensional space $\mathbb{R}^3$ is defined analogously.

Definition. The *2-sphere* (or *sphere*), is the surface S in $\mathbb{R}^3$ given by $S = \{(x, y, z) \in \mathbb{R}^3 \mid \sqrt{x^2 + y^2 + z^2} = r\}$ for some fixed radius r; we usually choose $r = 1$ for convenience.

A *transformation* is a mapping of one space to another that is one-to-one and onto; *i.e* it is a one-to-one correspondence from the set of points in one space into another [25]; transformations will be discussed more precisely in Section 2.1.2. Two spaces with the property that one can be continuously transformed into the other are said to be *topologically equivalent*.

The plane $\mathbb{R}^2$ is topologically equivalent to a sphere with radius $r = \infty$; and the surface of the (finite radius) sphere is topologically equivalent to a plane with points at infinity identified. Since all of the diagrams we consider in this work are finite, we can also consider the plane to be large but finite, and then the mapping between the plane and the sphere is more intuitive. Thus, to save much notational confusion it is useful to consider the following topological notions on the plane and assume that they can be generalized to the sphere where necessary; for example we will be rigorous about defining curves and subsets on the plane. Where the concepts in question are fundamentally different, for example the isometries of the plane as opposed to those of the sphere, we will define each in turn and the surface of their application will be evident from context.

Regarding notation, we will use Cartesian coordinates on the plane and in other-dimensional spaces except where noted, as there are situations in which radial (polar) coordinates make some ideas in 2- and 3-space much easier to work with. Rotations and angles, on the plane and 3-space, will usually be expressed in radians.

The sphere in n-dimensional space is topologically equivalent to a $(n-1)$-dimensional plane and so is often referred to as the $(n-1)$-*sphere*; the sphere in 3-space is properly referred to as a 2-sphere (and a circle is a 1-sphere), but we will usually just say "sphere" and use the more general term for higher-dimensional spheres.

Points on the plane are identified by their coordinates; two points with the same coordinate values are the same entity and are thus considered "identified". We thus have the notion of subsets of distinct points on the plane; when we consider symmetries we will see that a symmetry is an operation that identifies a subset, or collection of subsets, with itself. Our first fundamental topological notions are that of nearness of points, and connectedness of subsets. Let p be a point in the plane, and let A be some subset of the plane. We say that p is *near* A if every circular disk containing the point p also contains a point of A.

Definition. A subset S of the plane is *connected* if whenever S is divided into two non-empty disjoint subsets A and B ($S = A \cup B$), one of these sets always contains a point near the other.

Definition. A *curve* is the image of a continuous mapping from the closed interval $[0, 1]$ to the plane $\mathbb{R}^2$ (together with a point at infinity).

Furthermore, a curve is *closed* if the image of 0 and 1 are the same point, otherwise it is *open.* A *simple* closed curve has the property that its removal decomposes the plane into exactly two connected regions. Thus there is no $x \neq y \in [0, 1]$ such that x and y map onto the same point, *i.e* the mapping is one-to-one and the curve does not self-intersect.

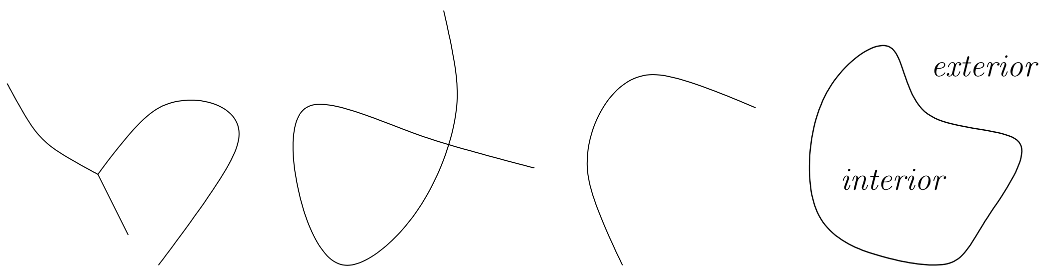
Definition. An *arc* is a curve that does not self-intersect and is not closed. An arc is called *non-trivial* if it is not a point.

A closed simple curve is the boundary of a bounded and simply connected set in the plane. Closed simple curves are often called *Jordan curve*s, or *pseudocircle*s by some authors (*i.e* [84]), since they can be continuously transformed (as we discuss later) to a circle. The Jordan curve theorem, a fundamental result in topology and the study of diagrams, states that the complement of a Jordan curve consists of two distinct connected regions, the (bounded) interior and the (unbounded) exterior. The Jordan curve theorem on the sphere still holds in the sense that a Jordan curve divides the sphere into two bounded connected regions.

The reader can assume that any curve that we discuss in this work is a simple closed Jordan curve, except where specifically noted. At times it is useful to specify the mapping from an interval to the plane which gives the curve; for example, on the plane indexed by the polar coordinates (θ, r), given $i \in [0, 1]$, the function $f(i) = (2\pi i, 1)$ gives a circle of radius one centred at the origin.

Definition. A collection of curves on a surface is referred to as a *diagram*.

As noted a curve decomposes the plane into two simply connected regions, one bounded and one unbounded (including the point at infinity): we refer to the bounded region as the curve's *interior* and the unbounded as the curve's *exterior*. Following common parlance we say that a curve *contains* any point or subset of points in its interior. In the case of the unit circle any point satisfying the equation $x^2 + y^2 < 1$ is contained in the interior. See Figure 2.1 for examples of the terminology.



(a) Not a curve (b) Not a curve due to self-intersection (c) An arc (d) Simple closed curve

Figure 2.1: Terminology for curves

Definition. A *region* in a diagram on the plane is a connected maximal subset of the plane that contains no points on a curve.

Let $\mathcal{D} = \{C_1, C_2, \ldots, C_n\}$ be a diagram composed of n curves in the plane. Every region is bounded except for one, the *exterior region*, which is composed of the n intersections of the curve exteriors. There are at least $n + 1$ regions present in any diagram, namely the interiors of all of the curves plus the exterior region, and more may occur if any of the curves intersect (share a point in common).

A common assumption in the literature is that the number of intersection points between curves is finite; in this work we always follow this assumption except for a unique diagram we present in Chapter 8.

Definition. A diagram is *simple* if no three curves intersect at a common point.

Some authors (eg. [115]) say that the curves are *in general position* if curves intersect transversally and the diagram is simple.

Given certain special properties of the diagram $\mathcal{C}$ we can refine our terminology.

Definition. A diagram $\mathcal{D}$ is an *independent family* if all of the 2^n sets given by the intersections $X_1 \cap X_2 \cap \cdots \cap X_n$ are non-empty, where each set X_i, $1 \leq i \leq n$, is the interior or the exterior of the curve C_i.

Note that these 2^n sets do not necessarily correspond to regions as some sets could be disconnected; see Figure 2.2, in which the set composed of the interiors of C_1 and C_2 includes two regions.

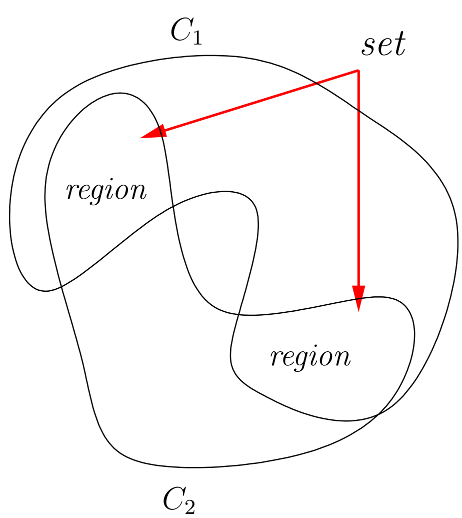


Figure 2.2: The difference between regions and sets; the intersection set $C_1 \cap C_2$ consists of two disconnected regions

Definition. An n-curve diagram $\mathcal{D}$ is an n*-Venn diagram*, or just *Venn diagram*, if $\mathcal{D}$ is an independent family and, additionally, each of the 2^n sets given by the intersections $X_1 \cap X_2 \cap \cdots \cap X_n$ is non-empty and connected and thus corresponds exactly to a region, where each set X_i, $1 \leq i \leq n$, is the interior or the exterior of the curve C_i.

Note that the term "Venn diagram", being the most commonly-used of the three types of diagram just defined, is often used informally to refer to diagrams with fewer or greater than 2^n regions, for example in using them for educational purposes to illustrate set inclusion and exclusion, but here we use them strictly as defined.

Definition. A diagram $\mathcal{D}$ is an *Euler diagram* if each of the 2^n sets given by the intersections $X_1 \cap X_2 \cap \cdots \cap X_n$ is connected, where each set X_i, $1 \leq i \leq n$, is the interior or the exterior of the curve C_i, and some of these intersections could be empty.

Table 2.1 sums up the differences, and shows the respective number of each regions in a diagram of each type of n curves.

	Euler diagram	**Venn diagram**	**independent family**
Sets	connected	connected and non-empty	non-empty
Regions	$\leq 2^n$	2^n	$\geq 2^n$

Table 2.1: Sets and regions for different types of diagrams

It is easy to see that in a simple Venn diagram curves must intersect transversally, but this may not be the case in independent families or in non-simple Venn diagrams.

Each of the 2^n regions in a diagram can be labelled (uniquely, in the case of Venn and Euler diagrams) by a subset of the n-set $\{1, 2, \ldots, n\}$, with the members of that subset corresponding exactly to the indices of the curves containing that region. Figure 2.3 shows the familiar three-curve Venn diagram with regions labelled by the indices of the curves containing them.

Definition. The *weight* of a region is the number of curves that contain it, and we can refer to a region of weight k, $0 \leq k \leq n$, as a k-region.

Two regions are called *adjacent* if their boundaries intersect in a non-trivial arc. A diagram is *monotone* if every region of weight k, $0 \leq k \leq n$, is adjacent to at least one region of weight $k+1$ (for $k < n$) and at least one region of weight $k-1$ (for $k > 0$). To illustrate, the diagram in Figure 2.3 is monotone, but see Figure 2.4 for a diagram that is not monotone: the region of weight one labelled $\{3\}$ is not adjacent to the region of weight zero.

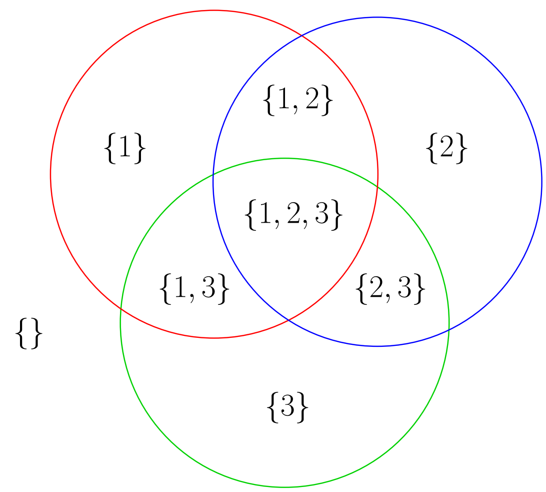


Figure 2.3: The (unique) simple 3-Venn diagram, with labelled regions

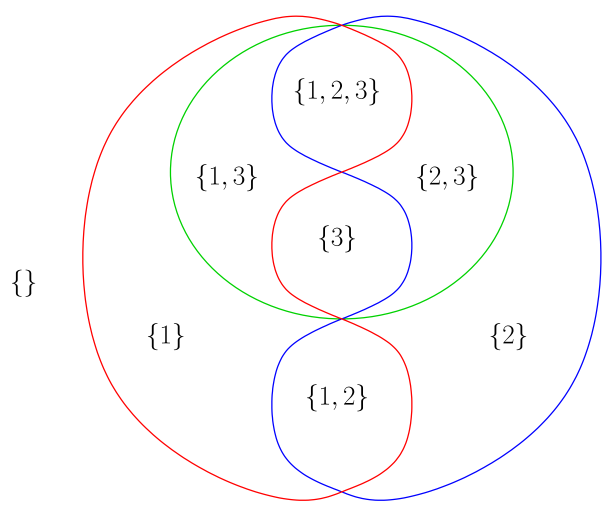


Figure 2.4: Example of a non-monotone 3-Venn diagram

2.1.1 History of Research in Diagrams

Now that we have established the important definitions and concepts, we will review the literature on the topic of diagrams before discussing transformations and more symmetry-related concepts.

As their name implies, Venn diagrams were formalized and studied extensively by John Venn [125], though the use of diagrams to illustrate set inclusion and exclusion predates him. Diagrammatic notations involving closed curves have been in use since at least the Middle Ages, for example by the Catalan polymath Llull (see [86], surveyed in [48]). The first known significant use of logical diagrams is by the renowned mathematical genius Euler [39] (whose work is recognized by having Euler diagrams named for him), and the German polymath Gottfried Wilhelm Leibniz [24, 23], who explored using lines, circles, and other pictorial representations of syllogisms (but see also [36] for an 11th-century example). It was Euler who popularized their use throughout the 19th century before Venn; as Baron [5] states, "through [Euler], knowledge of the diagrams became widespread and they had some considerable influence in the nineteenth century", influencing mathematicians such as Gergonne and Hamilton. The article [5] and especially the book [48] contain an historical context of logical diagrams, presenting some information on diagrammatic aids to reasoning dating from Aristotle.

The reader should note that the terminology between different types of diagrams has varied throughout the history of their study, most specifically with confusion between the terms "Euler" versus "Venn" diagrams. Often the term Venn diagram is used, if only informally, to refer to any diagram in which subsets correspond to connected regions. Confusion is especially prevalent in the use of diagrams to present logical arguments and set inclusion/exclusion information; many sources tend to describe an Euler diagram, or even an independent family, as a Venn diagram, when of course the set of Venn diagrams is the intersection of each of the other two kinds.

In [125], Venn showed the inadequacies of his contemporaries' approaches, and included a historical survey dating back to Euler. One of his more major contributions to the literature, in addition to formalizing various logical aspects of diagrams, is the first proof by construction of the existence of a Venn diagram for any number of curves. He also took great strides in clearing up terminological confusion by providing formal definitions for his diagrams, whereas Euler diagrams have not been formally studied in their own right until more recently: it is instructive to note that it was not until 2004 that a conference was organized specifically to bring together researchers in the area [102]. Research in Venn diagrams has traditionally appeared in computational geometry and discrete mathematics venues, as well as more educationally-oriented publications such as *Geombinatorics* and *Mathematical Gazette*.

Branko Grünbaum is rightly recognized as a pioneer in comprehensively studying various aspects of Venn diagrams, in a series of papers beginning with a prize-winning work from 1975 [56] in which he explored several geometric aspects of independent families and Venn diagrams, symmetric and otherwise. His further papers [41, 57, 58, 59] mostly discuss constructions of Venn diagrams, diagrams from different convex- and non-convex figures, and the combinatorics of counting sets and intersections.

Euler diagrams have been extensively used in the logic community for illustrating and proving syllogisms (for an example using Venn diagrams, see [55, ch. 3.3]); the reader is referred to comprehensive texts such as [78] or articles such as the entry for "Diagrams" in the Stanford Encyclopedia of Philosophy [118]. Dating from the work of Venn, Pierce [94] extended Venn diagrams to create a graphical system that was subsequently proven to be equivalent to a predicate language [101]. An important further extension was made by Shin [116, 117] in further increasing the expressive power of Pierce's representation system and placing it onto a formal foundation by proving its soundness. Other groups have provided other diagrammatic systems that are augmented in various ways [122]. Euler diagrams have recently been extensively studied by several groups in the UK as part of a project entitled "Reasoning with Diagrams". This research encompassed topics such as Euler diagram generation, counting, and syntax [42, 43] and understanding of diagrams based on aesthetic criteria [6]. Other projects have investigated the use of diagrams in automatic theorem proving [4, 76].

General combinatorial and topological properties of diagrams have been extensively studied by many researchers [31, 115], mostly from the standpoint of computational geometry. Independent families have been studied from a geometric standpoint as part of measure theory [40, 91].

Our historical survey of diagram research continues with a more comprehensive examination of the role of symmetry in diagrams in Chapter 3.

2.1.2 Transformations of Curves

We must now establish, via some careful definitions, when two diagrams on the plane are similar or different in some sense.

Recall the concept of nearness from Section 2.1 earlier; nearness lets us define transformations on the plane that can turn one diagram into another without changing essential properties of the diagrams.

Definition. A *continuous transformation* from one subset D of the plane to another subset R is a function f with domain D and range R such that for any point $p \in D$ and set $A \subseteq D$, if p is near A, then $f(p)$ is near the set $f(A) = \{f(q)|q \in A\}$.

A curve as we have defined it is any subset of the plane that can be continuously transformed to a circle, and a region of a diagram is a subset that can be continuously transformed to an open disc (for example, the subset of points satisfying $x^2 + y^2 < 1$, called the *(open) unit disc*).

Intuitively we can think of continuous transformations by imagining that the plane

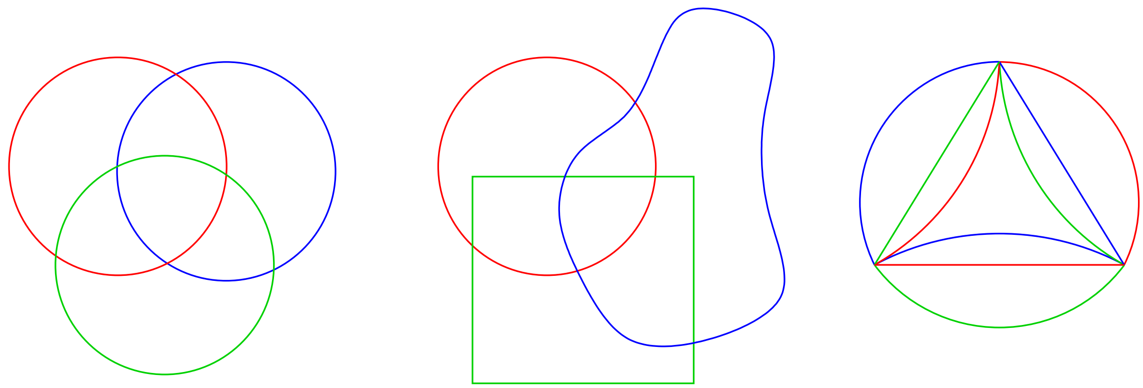


Figure 2.5: Two isomorphic diagrams, on the left, and a third that is not isomorphic to either of the first two

is a sheet of rubber that we can stretch and distort in various ways, but without folding or breaking the rubber. Simple transformations are stretching, shrinking, or distorting in various directions, along with translations, rotations and reflections. We shall only be concerned with continuous transformations on diagrams which preserve Jordan curves.

We use the concept of a *permutation* of an ordered list S: a permutation of S is a rearrangement of the elements of S into a one-to-one correspondence with S itself (permutations are further discussed in Section 2.3.4).

Definition. Two diagrams $\mathcal{C} = \{C_1, C_2, \ldots, C_n\}$ and $\mathcal{D} = \{D_1, D_2, \ldots, D_n\}$ are *isomorphic* if there exists a continuous transformation f from the surface to itself and a permutation π of $\{1, 2 \ldots, n\}$ such that $f(C_i) = D_{\pi(j)}$ for all $1 \leq i \leq n$.

For example, a transformation f on the plane is $f : \mathbb{R}^2 \to \mathbb{R}^2$, and on the 2-sphere S (as defined earlier) is $f : S \to S$.

Continuous transformations are invertible and transitive (see for example [73]), so the isomorphism relation is an equivalence relation. See Figure 2.5 for an example of isomorphism between diagrams; the first two diagrams, on the left, can clearly be transformed to each other by continuous transformations, but the third has a different vertex and edge count and thus is nonisomorphic to the others.

2.1.3 Isometries and Congruence

We now need to specify when two curves in a given diagram are in some sense the same "shape". We say that a *symmetry* is any transformation on an object under which the object remains invariant in some sense; this definition will be refined in future chapters, when we consider colour symmetry. We are interested in symmetries that preserve geometrical properties such as distance, length, and area. To measure the distance between two points, on the plane we use Euclidean distance, and on the sphere we use the shortest distance between two points travelling on the surface of the sphere—such a distance follows a geodesic (a circle whose centre is coincident with the centre of the sphere) [104].

Definition. An *isometry* on a surface is a distance-preserving continuous transformation. More precisely, a continuous transformation f on a surface is an isometry if, for every two points p, q with distance $d(p, q)$, it is always true that $d(f(p), f(q)) = d(p, q)$.

The word isometry comes the Greek and literally means "equal measure". The isometries of the plane of Euclidean geometry are translations, reflections, rotations, and glide reflections:

translation : a translation T_v of a point $p = (x, y)$ shifts p in the direction of a vector $\vec{v} = (v_x, v_y)$; that is, $T_{\vec{v}}(p) = (x + v_x, y + v_y)$.

rotation : a rotation $R_{c,\theta}$ of $p = (x, y)$ is a rotation by angle θ, $0 \leq \theta < 2\pi$, about the centre of rotation c. Using polar coordinates to rotate the point $p = (r, \phi)$, if c is the origin at $(0, 0)$, $R_{c,\theta}(p) = (r, \phi + \theta)$. If c is not the origin, $R_{c,\theta}$ can be performed by performing a translation mapping c to the origin, performing $R_{(0,0),\theta}$, and then translating the origin back to c. We also note that rotations can be characterized as those isometries that fix exactly one point, namely c.

reflection : a reflection is denoted $F_{c,\vec{v}}$, where $c = (c_x, c_y)$ is a point and $\vec{v} = (v_x, v_y)$ a (unit) vector; a point $p = (x, y)$ is reflected across the line L perpendicular to $\vec{v}$ passing through c. The formula is given by first finding the component of $p - c$ in the v direction, and then subtracting that twice from p, so

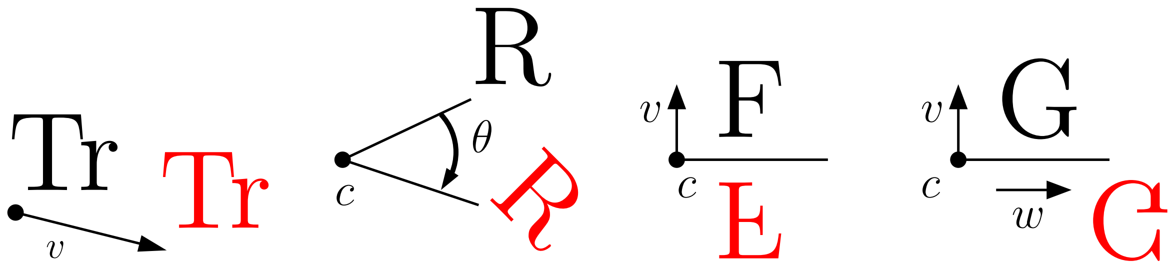
$$F_{c,\vec{v}}(p) = (x, y) - 2((x - c_x)v_x, (y - c_y)v_y) = (x - 2(x - c_x)v_x, y - 2(y - c_y)v_y)\,.$$

glide reflection : the fourth type of isometry is a composition of reflection and translation. A glide reflection $G_{c,\vec{w}}$ of a point p is a reflection in the line through c containing the vector $\vec{w}$ followed by translation in the direction of vector $\vec{w}$. That is, $G_{c,\vec{w}}(p) = \vec{w} + F_{c,\vec{v}}(p)$, where $\vec{v}$ is perpendicular to $\vec{w}$, and it is also true that $G_{c,\vec{w}}(p) = F_{c,\vec{v}}(p + \vec{w})$, that is, it does not matter in which order the reflection and translation are composed.

It has been shown (see, for example [25]) that these exhaust the possible isometries of the plane, and it can be appreciated by considering what transformation results by composing two or more of these operations. For example, the composition of two translations is a translation, the composition of a translation and a rotation is another rotation, and a translation can be expressed as the composition of two rotations. We follow most authors by including reflection as a separate isometry from glide reflection due to its simplicity, though a reflection can be regarded as a glide reflection with a trivial translation of zero. Furthermore, any translation or rotation can be expressed as the composition of two reflections; thus, all plane isometries can be generated just by glide reflections. See Figure 2.6 for visual examples of all of the isometries considered.

A trivial translation or rotation of distance or angle zero (respectively) is often regarded as a separate isometry itself:

identity : the identity function $I(p) = p$.



(a) Translation by $\vec{v}$

(b) Rotation by θ around c

(c) Reflection by $\vec{v}$ from c

(d) Glide reflection along $\vec{w}$ by c

Figure 2.6: Isometries in the plane

The isometries of the sphere will be discussed in Section 4.2.

Finally, we note that there can be functions on a diagram, such as dilations, that show evidence of self-similarity but are not isometries; in this work we only consider symmetries that are isometries, except in a few specific cases.

For the following definition, we wish to not allow reflections in moving one curve onto another: the two remaining transformations, translations and rotations (and compositions of them), are called *direct isometries*.

Definition. Two curves C_1 and C_2 in a diagram are *congruent* if there is some direct isometry of the plane $f : \mathbb{R}^2 \to \mathbb{R}^2$ such that $f(C_1) = C_2$.

See Figure 2.7 for an example. The notion of congruence generalizes naturally to curves on any surface. Note that the isometry involved must be direct: for example, two curves that differ only by a reflection are not considered congruent, though some authors ([58, 106]) consider the notion of congruence to include reflection.

2.2 Graphs

It will be helpful to define some basic graph terms as often diagrams are referred to in a graph-theoretic sense. We follow West [130] in our basic terminology.

Definition. A *graph* $G = (V, E)$ is a set of vertices $V = \{v_1, v_2, \ldots, v_n\}$ and a set of edges $E = \{e_1, e_2, \ldots, e_m\}$ where each edge is an unordered pair of vertices

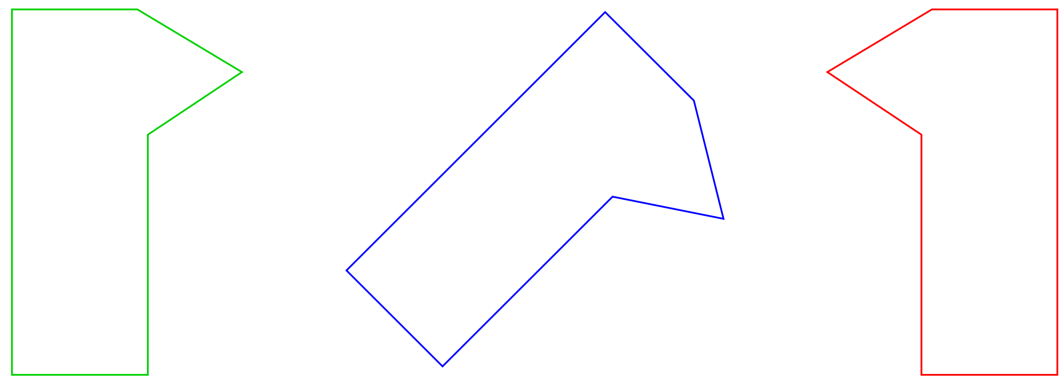


Figure 2.7: Two congruent curves, on the left, and a third that is not congruent to either of the first two

$e_i = \{v_j, v_k\} \subseteq E$, for $1 \leq i \leq m$ and $1 \leq j \neq k \leq n$, and vertices v_j and v_k are then termed *adjacent*, since there is an edge (e_i) between them.

The number of edges sharing an endpoint v is called the *degree* of v; a graph G is called k-*regular* if all $v \in V(G)$ have degree k. For a given graph G the vertex set V and the edge set E can be written $V(G)$ and $E(G)$. A graph H is a *subgraph* of G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$.

Given an edge $e_i = \{v_j, v_k\}$, the vertices v_j and v_k are called *endpoints* of that edge, and the edge is *incident* to its endpoints. Given a graph G and two vertices $v_j, v_k \in V(G)$, there may exist multiple edges of the form $\{v_j, v_k\} \in E(G)$, which are termed *multiedges*. An edge of the form $\{v_j, v_j\}$ is called a *loop*. A graph without multiedges or loops is termed *simple*, but we shall avoid this term in general since it also refers to a concept specific to diagrams.

It is often very handy to associate some information with the edges and/or vertices in a graph. A graph is *edge-labelled* or *edge-coloured* if each edge has a label associated with it, where a label can be any character string. Thus an *edge-labelling*, or *edge-colouring* of a graph G is a function $label(G) : E(G) \rightarrow L$ from the set of edges onto a (finite) set of labels L; so every edge will receive a (not necessarily distinct) label. A graph can be vertex-labelled in a similar fashion. Common label sets are some set of colours, *e.g.* $L = \{\text{red}, \text{black}\}$, or $L = \mathbb{N}$, the natural numbers. We will often refer to either an edge- or vertex-coloured graph as just a *labelled* or *coloured* graph, where the set of labels and whether the edges, vertices, or both are labelled is clear from context; often the term *colour* refers to edge-labelling. The full definition of a graph is thus $G = (V, E, \mathit{label})$, where *label* is the labelling function. Note that *label* could be trivial for an unlabelled graph, *i.e* $\mathit{label}(v) = \mathit{label}(e) = \mathit{null}$, for all $v, e \in G$ and some irrelevant *null* value. Thus we will assume that all graphs are edge- and vertex-labelled, and the labelling function is always implicitly present in the graph definition. Often two different labellings are the same in some sense. Two labellings of a graph G, using the same label set, are called *equivalent* if one can be transformed into the other by applying a permutation to the labels; this reflects the fact that the curves in a diagram are traditionally numbered from 1 to n but the order is usually irrelevant.

A *path* in a graph, from vertices v_1 to v_k, is a sequence of edges $(v_1, v_2),\ (v_2, v_3),\ \ldots\ ,$ (v_{k-1}, v_k) such that $v_i \neq v_j$ for $i \neq j$. A graph is *connected* if there is a path between any pair of distinct vertices, and *disconnected* otherwise. A graph is k*-connected* if the removal of any set of up to $k-1$ vertices does not separate it into disconnected parts.

2.2.1 Embeddings and Diagrams

When working on the plane, we can represent graphs with vertices as points and edges as lines between points. Recall that an arc is defined to be the image of a continuous one-to-one map from $[0, 1]$ to the plane $\mathbb{R}^2$ where no $x \neq y$ map to the same point; thus an arc is very similar to a curve except that the ends do not necessarily meet. An edge of a graph can thus be represented by an arc.

Definition. An *embedding* of a graph is a function f defined on $V(G) \cup E(G)$ that assigns each unique vertex v to a unique point $f(v)$ in the plane and assigns each edge with endpoints v_i, v_j to an arc with endpoints at $f(v_i)$ and $f(v_j)$.

A graph G is *planar* if there is an embedding of it such that its vertices map to distinct points and its edges map to distinct arcs, such that no point in the plane is shared between two or more arcs unless that point is also a vertex. A *plane graph* is a graph G together with a particular embedding of G.

Embedding a graph in a planar fashion gives us extra structure that we can use. A *face* of a plane graph is a maximal region of the plane containing no points on an edge or vertex. Faces are all open bounded regions except for one, the unbounded external face.

A diagram can be represented in the obvious way as a plane graph, by treating curve intersections as vertices and segments of curves between intersections as labelled edges, with an edge e labelled with the index i of the curve C_i it is part of. We noted that the regions of an n-curve diagram can be labelled by the unique subset of $\{1, 2, \ldots, n\}$ of the curves that contain it; this label can also be applied to the faces of the plane graph in the same fashion. The edges of the plane graph are labelled with the index, or colour, of the corresponding curve that the edge is a subset of. We can overload the terms "diagram" (and also "Euler diagram", "Venn diagram", and "independent family") to also refer to this labelled graph corresponding to a diagram; it will be clear from context which representation of the diagram we are using, and we can always switch between them at will.

We must be careful to note that a diagram, when represented as a plane graph, gives exactly one plane graph but one graph may correspond to several diagrams, depending on how its components are embedded in the plane. We will see many examples of diagrams that have the same underlying graph structure but are different diagrams since they have different embeddings in the plane.

Recall that a diagram is *simple* if no three curves intersect in a common point, and all curves intersect transversally. A simple diagram is thus 4-regular, since every vertex must have exactly four incident edges, two from each of the curves that cross at it. In a simple Venn diagram, since every vertex has degree four, Euler's formula ($V - E + F = 2$), combined with the fact that there are 2^n faces, tells us that there are $2^n - 2$ vertices and $2^{n+1} - 4$ edges in a simple Venn diagram.

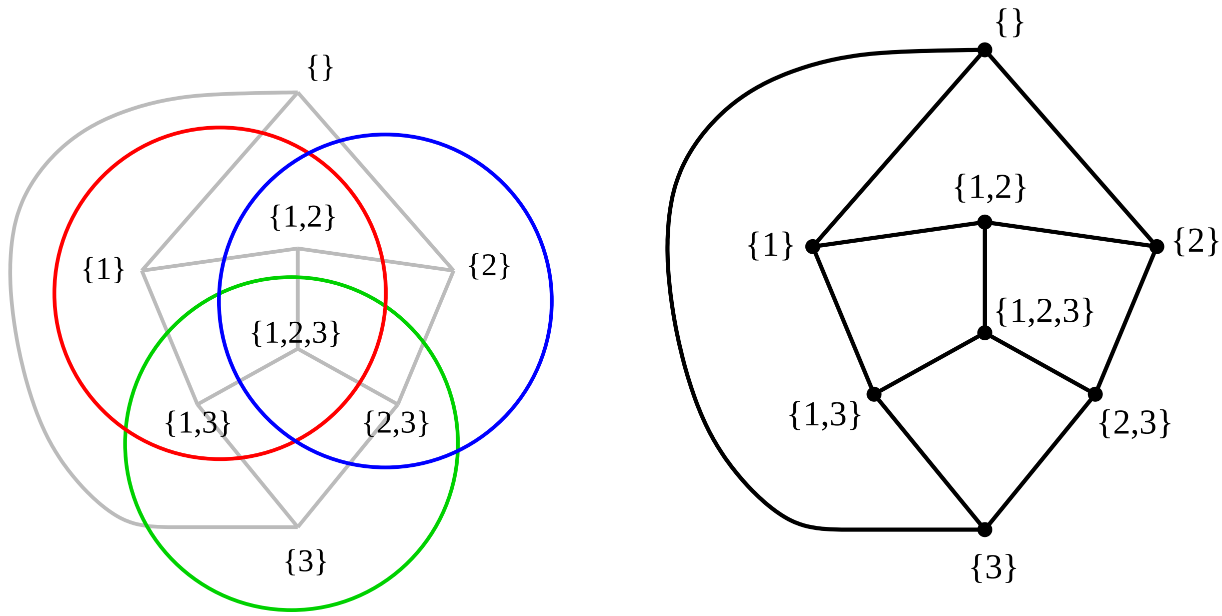
2.2.2 Dual Graphs

From any plane graph we can form a related plane graph containing much the same information.

Definition ([130])**.** The *dual graph* G^* of a plane graph G is a plane graph whose vertices correspond to the faces of G. The edges of G^* correspond to the edges of G as follows: if e is an edge of G with face X on one side and face Y on the other side, then the endpoints of the dual edge $e^* \in E(G^*)$ are the vertices x, y that represent the faces X, Y of G. The clockwise order in the plane of the edges incident of $x \in V(G^*)$ is the order of the edges bounding the face X of G in a clockwise walk around its boundary.

When drawing the dual graph G^* it is usual to place a vertex $v^* \in V(G^*)$ inside the interior of the face it corresponds to in G, and to draw each dual edge e^* such that it crosses its corresponding edge $e \in G$. Hence G^* is also a plane graph, and each edge in $E(G^*)$ in this layout crosses exactly one edge of G. The dual graph of the dual graph can be drawn to be exactly concurrent to G, so we can view a plane graph G and its dual as two aspects of the same structure and consideration of one can give insights about the other.

If a graph G is a labelled graph, an edge e^* in the dual graph G^* can be labelled with the label of its corresponding edge $e \in E(G)$. If the faces in G are labelled, a vertex $v \in V(G^*)$ can be labelled with the label of the face it corresponds to. We will often deal with diagrams and their duals, which are usually labelled in this way. See Figure 2.8 for an example of a labelled diagram and its labelled dual.



(a) The diagram, with labelled faces, and its dual (b) The labelled dual

Figure 2.8: The 3-Venn diagram and its labelled dual

We usually deal with diagrams with the property that all curves intersect at least one other curve (and thus there are at least two curves); a curve that does not intersect any other curve is called *isolated*. A diagram with no isolated curves where curves intersect transversally is a two-connected plane graph, and its dual is also two-connected [16]; recall that multiple edges are allowed between two vertices in a graph. Furthermore, a simple diagram with no isolated curves is a 4-regular graph. Thus, all of the faces of the dual of a diagram are 4-faces (faces bordered by exactly 4 edges) if and only if the corresponding diagram is simple. These and many more graph-theoretic properties of diagrams and their duals are discussed in [16, 17, 18]. Throughout this work we assume any diagram under discussion has no isolated curves, unless otherwise specified: for example the 1-curve Venn diagram, consisting of a single (isolated) curve.

2.3 Strings and Posets

In this section we discuss some properties of sets of strings that will prove useful. We follow Trotter [123, 124] for our discussion of posets.

Recall from Section 2.1 that each of the 2^n regions in a Venn diagram can be uniquely labelled by a subset of $\{1, 2, \ldots, n\}$, with the members of that subset corresponding exactly to the indices of the curves containing that region. We often find it convenient to refer to a region's label as a string of bits (or *bitstring*) of length n in which the bit at position i is set to 1 if and only if curve i contains that region, and 0 otherwise. Figure 2.9 shows the bitstring labelling of the dual graph from Figure 2.8. The weight of the region is thus the number of 1's in the corresponding bitstring. The weight of a bitstring x is written $|x|$.

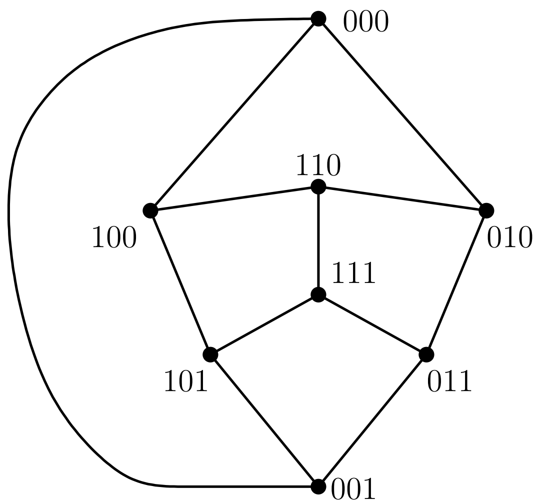


Figure 2.9: Bitstring labelling of dual graph from Figure 2.8

Let $B_n = \{0, 1\}^n$ be the set of 2^n unique bitstrings of length n. B_n thus has an exact correspondence with the set of 2^n subsets of the n-set, which is referred to as the *power set* of the n-set, and written $2^{[n]}$.

2.3.1 Posets and the Boolean Lattice

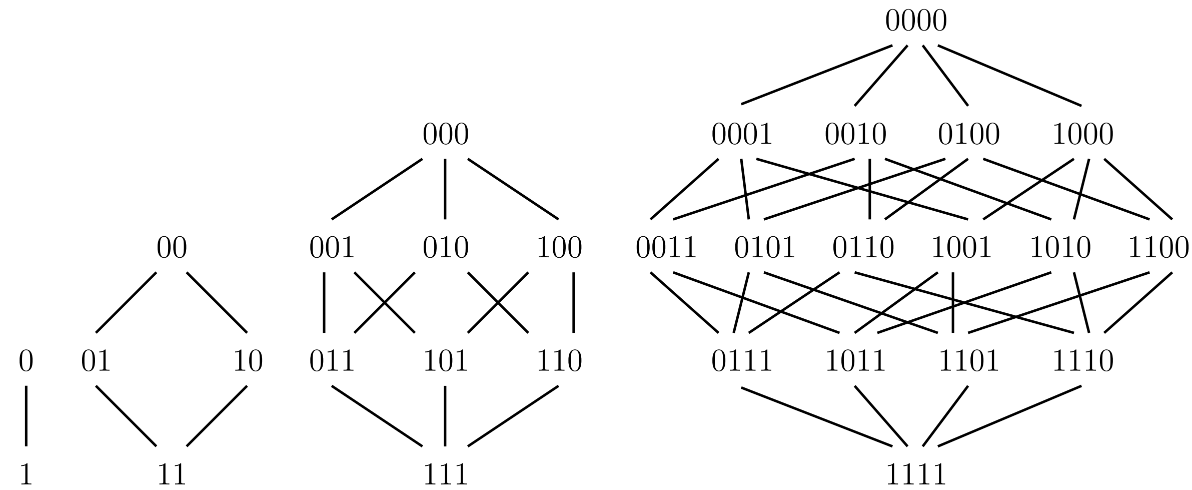
Definition. A *partially ordered set* or *poset* is a pair (X, P) where X is a set and P is a binary relation on X. If two elements $x, y \in X$ are related (*i.e* $(x, y) \in P$) we write $x \prec y$, with the sets X and the relation P understood from context. For (X, P) to be a poset, P must be

- *reflexive*: $a \prec a$ for all $a \in X$,
- *antisymmetric*: if $a \prec b$ then $b \not\prec a$ for all distinct $a, b \in X$,
- and *transitive*: if $a \prec b$ and $b \prec c$, then $a \prec c$, for all distinct $a, b, c \in X$.

Two elements are called *comparable* if there exists a relation between them, and they are *incomparable* otherwise. A *cover relation* $a \prec b$ is a relation for which there does not exist $c \in X$ such that $a \prec c \prec b$; we can say that "a is covered by b" .

A *Hasse diagram* is a representation of a poset as a graph together with an embedding in the plane; nodes are the elements of X and edges connect those elements that are related by P. For clarity and simplicity, transitive relations are not represented: thus only cover relations are drawn as edges in the Hasse diagram. Elements are laid out in the plane according to P according to the convention that if $a \prec b$ then a is drawn higher (in a position with higher y-coordinate, given a vertical y-axis on the plane with coordinates increasing upward).

In this work we will be dealing almost exclusively with the poset called the *boolean* (or *subset*) *lattice of order n*. In the boolean lattice, the elements are the set B_n, corresponding in the way described above to the 2^n subsets of the n-set, and they are related by subset inclusion; the boolean lattice is written $\mathcal{B}_n$, where $\mathcal{B}_n = (B_n, \subseteq)$. Two bitstrings a, b are related (and we also say $a \prec b$) if and only if their corresponding subsets are related such that $a \subset b$. Thus, the cover relations are exactly those pairs of bitstrings a, b such that a and b differ in one bit position where a has



(a) $\mathcal{B}_1$ (b) $\mathcal{B}_2$ (c) $\mathcal{B}_3$ (d) $\mathcal{B}_4$

Figure 2.10: Hasse diagrams of small boolean lattices

a 0 and b has a 1. In the Hasse diagram of the boolean lattice, the all-zero bitstring 0^n, corresponding to the empty set, appears as the highest element, and the all-one bitstring 1^n as the lowest. Figure 2.10 shows the Hasse diagrams of the boolean lattices of small orders.

2.3.2 Chains in Posets

Definition. A *chain* is a set of elements, each pair of which is comparable.

The maximum length of a chain in the boolean lattice $\mathcal{B}^n$ is $n+1$ since each element must have weight at least one greater than those preceding it and at least one less than those following it. A chain C of a poset P is *maximal* if no element can be added to it. Furthermore, C is called *saturated* if there does not exist $b \in P - C$ such that $a \prec b \prec c$ for some $a, c \in C$, and such that $C \cup \{b\}$ is a chain. In this work all chains are saturated unless otherwise specified. We will always write a chain's elements in the order $\{\ \sigma_1,\ \sigma_2,\ \ldots\ , \sigma_k\ \}$, where $\sigma_i\ \prec\ \sigma_j$ for $i < j$, and the braces can be omitted for brevity.

The *height* of a poset P is the largest h for which there exists a chain of length h in P. The boolean lattice of order n clearly has height $n+1$.

A poset is said to be *ranked* if all maximal chains have the same length; in the case of the boolean lattice all maximal chains have length $n+1$. Thus, the poset can be partitioned into ranks, where every maximal chain consists of exactly one element from each rank. The ranks can be indexed by their distance (in the graph-theoretic sense) in the Hasse diagram from an element with the lowest height; in the boolean lattice the natural rank for a bitstring is its weight, which is its distance from the extremal element 0^n.

In a ranked poset, two elements of the same rank can never be in the same chain. A set of incomparable elements is called an *antichain*, and it is maximal if no elements can be added to it. The *width* of a poset P is the largest w for which there exists an antichain of size w in P. The set of bitstrings of weight k form a maximal antichain in $\mathcal{B}_n$ for $0 \le k \le n$. A famous theorem in poset theory due to Sperner [121] asserts that no antichain in $\mathcal{B}_n$ can have more than $\binom{n}{\lfloor n/2 \rfloor}$ elements, which is the middle binomial coefficient, or the number of bitstrings with $\lfloor n/2 \rfloor$ of their bits set to 1. Thus the width of the boolean lattice $\mathcal{B}_n$ is $\binom{n}{\lfloor n/2 \rfloor}$.

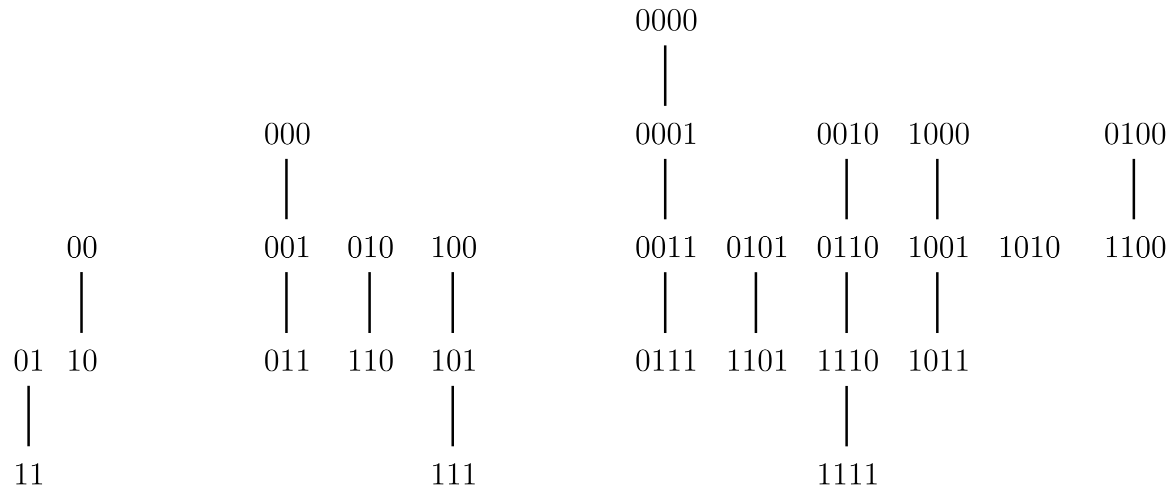
2.3.3 Chain Decompositions

Dilworth's Theorem [29] asserts that a poset of width w can be partitioned into w disjoint chains, and there is no partition into fewer chains; combined with Sperner's Theorem this tells us that $\mathcal{B}_n$ can be partitioned into $\binom{n}{\lfloor n/2 \rfloor}$ chains.

Definition. A partition of a poset into chains is called a *chain decomposition.*

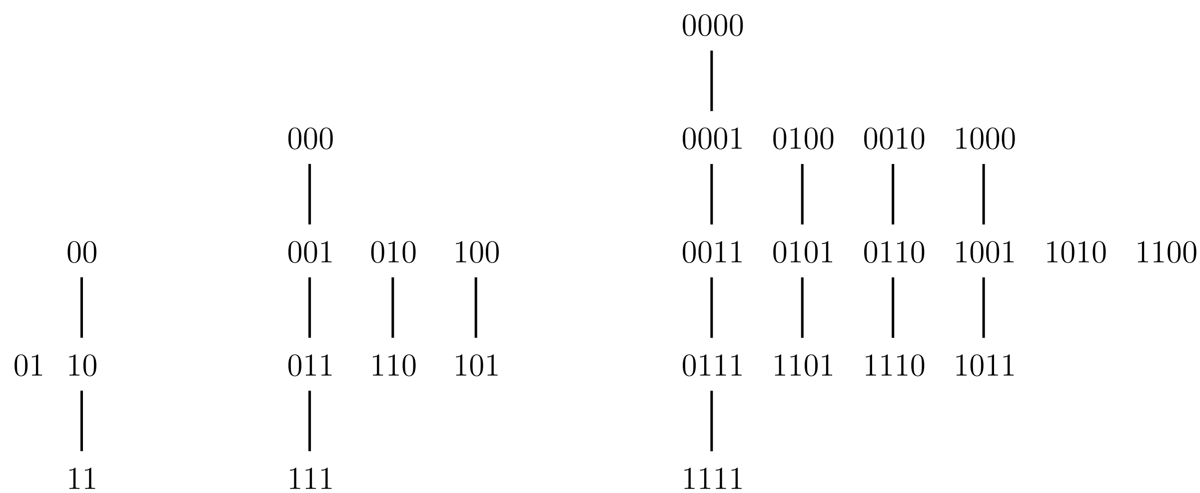
Figure 2.11 shows some chain decompositions of the posets in Figure 2.10. The entire Hasse diagram for $\mathcal{B}_1$ is a single chain, so Figure 2.10(a) is clearly a chain decomposition.

A chain $C = \{\sigma_1, \sigma_2, \ldots, \sigma_k\}$ in a ranked poset P of height h is called a *symmetric* chain if there exists an integer i such that C contains exactly one element from each rank $A_i, A_{i+1}, \ldots, A_{h+1-i}$. A symmetric chain is thus "balanced" about the middle



(a) $\mathcal{B}_2$ (b) $\mathcal{B}_3$ (c) $\mathcal{B}_4$

Figure 2.11: Chain decompositions of the boolean lattices in Figure 2.10



(a) $\mathcal{B}_2$ (b) $\mathcal{B}_3$ (c) $\mathcal{B}_4$

Figure 2.12: Symmetric chain decompositions of the boolean lattices in Figure 2.10

rank of the poset and saturated. A *symmetric chain decomposition* is a partition of a poset into symmetric saturated chains; for the boolean lattice there are thus $\binom{n}{\lfloor n/2 \rfloor}$ chains in a symmetric chain decomposition. The chain decompositions in Figure 2.11 are not symmetric (though note that the unique decomposition of $\mathcal{B}_1$ is symmetric); some symmetric chain decompositions of the small order lattices are shown in Figure 2.12.

A very nice chain decomposition for the boolean lattice $\mathcal{B}_n$ was given by de Bruijn, et al. in 1951 [28], and studied by many subsequent authors; see, for example [70] and [132]. Their construction uses the theory of balanced strings of parentheses, which are deeply related to binary trees and Catalan numbers, both fundamental topics in computer science. A properly nested binary string corresponds exactly to a balanced parentheses string with 1s representing left parentheses '(' and 0s right parentheses ')'; that is, every 1 in the binary string has a matching 0 that is the closest 0 that is otherwise unmatched to its right, and similarly every 0 has a matching 1 to its left; for example

$$1\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 0$$

is a string with six matched pairs.

Greene and Kleitman [53] (see also [1]) showed that de Bruijn's method of producing the decomposition is equivalent to the following. Any bitstring in $\mathcal{B}_n$ of length n can be uniquely written in the form

$$\alpha_0\, 0\ \ldots\ \alpha_{p-1}\, 0\, \alpha_p\, 1\, \alpha_{p+1}\ \ldots\ 1\, \alpha_q \tag{2.1}$$

for some p and q with $0 \leq p \leq q$, where each substring α_i is a properly nested string, and possibly empty; there are exactly p 0s and $q - p$ 1s that are "free" in the sense that they have no match. The string 2.1 is part of a chain of length $q + 1$,

$$\alpha_0\, 0\ \ldots\ \alpha_{q-1}\, 0\, \alpha_q,\ \ \alpha_0\, 0\ \ldots\ \alpha_{q-2}\, 0\, \alpha_{q-1}\, 1\, \alpha_q,\ \cdots\ ,\ \alpha_0\, 1\, \alpha_1\ \ldots\ 1\, \alpha_q\,,$$

in which we start with q free 0s and change them, from right to left, into free 1s. It is easy to see that every bitstring of length n falls into exactly one chain, so the decomposition is indeed a chain decomposition, and the chains so produced are symmetric and maximal, and thus the resulting partition is a symmetric chain decomposition.

Note especially that each chain begins with a string with no unmatched 1s and finishes with a string with no unmatched 0s.

Knuth [83] presents the de Bruijn chain decomposition as a recursive construction as follows. The base case is the decomposition for $n = 1$, which is the single chain $\{0, 1\}$. Given a chain decomposition of the order-n lattice, to create the chain decomposition for the order-$(n+1)$ lattice, take every every chain $\{\sigma_k, \sigma_{k+1}, \ldots, \sigma_{k+j}\}$, where σ_i is a bitstring of weight i, and replace it by the two chains

$$\{\ \sigma_{k+1}\,0,\ \ldots\ ,\ \sigma_{k+j}\,0\ \},\quad \text{and}$$
$$\{\ \sigma_k\,0,\ \sigma_k\,1,\ \ldots\ ,\ \sigma_{k+j-1}\,1,\ \sigma_{k+j}\,1\ \},$$

and the first of these chains is omitted when $j = 0$ (so there is only one element in the chain in the order-n lattice).

Greene and Kleitman [53] showed that the properly-nested strings approach and the recursive construction give the same chains.

In Chapters 3 and 5 we will discuss further properties of this chain decomposition as used to create monotone Venn diagrams and modifications of it to create symmetric Venn diagrams on the sphere.

2.3.4 Permutations

Permutations are a fundamental combinatorial object that we use often in discussing how a symmetry will permute the curve colours of a diagram. Given the n-set $[n] = \{1, 2, ..., n\}$ a *permutation* is an arrangement of the elements of $[n]$ in some order; clearly there are $n!$ distinct such permutations. A permutation π of $1\ 2\ \ldots\ n$ can be written as $\pi(1)\ \pi(2)\ \ldots\ \pi(n)$; this is often referred to as "one-line" notation. For example, the permutation of the 5-set that shifts the final three elements 3 4 5 to 5 3 4 can be written 12534. The set of permutations of the n-set is often written $\mathbb{P}_n$.

It is often easier to write a permutation in *cycle notation*, which compactly displays which elements map onto other elements. A k-*cycle* in a permutation is a sequence of distinct elements $x_1\ x_2\ \ldots\ x_k$ such that $x_i = \pi(x_{i-1})$ for $2 \leq i \leq k$ and $x_1 = \pi(x_k)$. Any permutation can be written as a product of distinct cycles, so our example 12534 could be written (1)(2)(354).

A permutation is *circular* if it consists of one n-cycle (see [52, pp. 259–262]).

2.4 Group Theory

In this section we discuss some basic group theory essential to our understanding of symmetries. We follow standard notation as much as possible. A good reference is Dummit and Foote [30]; an entertaining introduction with many geometric examples is Budden [10].

Definition. A *group* $(\mathsf{G}, \circ)$ is a set G with a binary operation $\circ$ which satisfies the following axioms:

- The set is closed under the operation: for every $a, b \in \mathsf{G}$, $a \circ b \in \mathsf{G}$.
- The operation is associative: for every $a, b, c \in \mathsf{G}$, $a \circ (b \circ c) = (a \circ b) \circ c$.
- There exists an identity: there is an element $e \in \mathsf{G}$ such that $e \circ a = a \circ e = a$ for every $a \in \mathsf{G}$.
- There exist inverses: for every $a \in \mathsf{G}$ there is an element $b \in \mathsf{G}$ (called the *inverse* of a, written a^{-1}) such that $a \circ b = b \circ a = e$.

The number of elements in G, written as $|\mathsf{G}|$, is referred to as the *order* of G. We are often more informal and just write that "G is a group under $\circ$", or even "G is a group" if the operation $\circ$ is implied or assumed from context. A group is *Abelian* (or *commutative*) if the operation $\circ$ is commutative (that is, for every $a, b \in \mathsf{G}$, $a \circ b = b \circ a$).

An important concept is that of subgroups.

Definition. Let G be a group under the operation $\circ$. A subset H of G is a *subgroup* of G, written $\mathsf{H} < \mathsf{G}$, if H is nonempty and $x, y \in \mathsf{H}$ implies $x^{-1} \in \mathsf{H}$ and $xy \in \mathsf{H}$.

Lagrange's theorem states that, if $\mathsf{H} < \mathsf{G}$, then $|\mathsf{H}| \,/\, |\mathsf{G}|$: the order of H must divide the order of G.

Finally, we need to express when two different groups are really the same.

Definition. An *isomorphism* from a group $\mathsf{G} = (X, \circ)$ to a group $\mathsf{H} = (Y, \star)$ is a one-to-one mapping (or function) $f : \mathsf{G} \to \mathsf{H}$ that preserves the group operation. That is, for all $a, b \in \mathsf{X}$, $f(a \circ b) = f(a) \star f(b)$.

If there is an isomorphism from G onto H, we say that G and H are *isomorphic*, and we write $\mathsf{G} \cong \mathsf{H}$.

Simple examples of groups include the integers $\mathbb{Z}$ under addition, the group of translations of the plane (the elements are translations and the binary operation is "followed by"), or the set of permutations of a set of n elements where the operation is composition of permutations. As an example of a subgroup, the even integers under addition form a subgroup of $\mathbb{Z}$ under addition—note that the odd integers do not, since addition of odd integers does not give another odd integer.

The number and properties of small groups of finite size are well-studied, and the literature on the subject is vast. A common source of confusion is the notation distinguishing various groups, since different notations emphasize different properties and have been in vogue for different mathematicians at different periods. In later sections we will present the notation we use for particular groups, which is a simplification of several common notation schemes.

2.4.1 Groups of Diagrams

It is important to distinguish the notions of a group and the object it is acting on. A *group action* of a group G onto a set A is a function from $\mathsf{G} \times A$ to A (with some additional properties). In our case, the set A is a diagram and G is the group of isometries that map the diagram back onto itself (that the set of isometries of a diagram form a group is well-known [10]).

Definition. Let D be a diagram on a surface. The group formed by the set of isometries of D, along with the operation of composition of isometries, is termed the *symmetry group* of D (not to be confused with the *symmetric group*, defined below, which is a particular abstract group).

Given some diagram D, its symmetry group is generally written $S(D)$, and we say that the group $S(D)$ *acts on* D. When we are constructing a specific diagram in order to show that its symmetry group is isomorphic to a known group, we say that the diagram *realizes* the group in question. This follows standard terminology ([10]).

2.4.2 Generators, Orbits, and Fundamental Domains

An important aspect of groups acting on geometric objects is choosing a set of elements from the set being acted upon that, under the group's operations, can give the entire set. A group acting on a set partitions the set into disjoint equivalence classes under the action of G; these classes are called orbits. The orbit of an element x being acted upon is the set of elements to which x can be taken by the elements of G.

Definition. Let G be a group acting on the set A and let $x \in A$. The equivalence class $\{g(x) \mid g \in \mathsf{G}\}$ is called the *orbit* of x.

It is useful to have a set of representatives of the group that can generate all of the group elements.

Definition. Let G be a group, and $S \subseteq \mathsf{G}$ be a subset of the elements of G. Then S is a set of *generators*, or *generating set*, if every element of G can be written as a finite product of the elements of S and their inverses. Informally we say that S *generates* G.

For example, the set $\{1\}$ is a set of generators for the additive group of integers $\mathbb{Z}$ since every integer is a sum of a finite number of 1s and its inverse (-1). In our examples of rotationally symmetric diagrams of n curves, the operation of a clockwise rotation by $2\pi/n$ about the point of rotation is a generator, since all possible rotations can be described by one or more application of this single rotation. Each curve belongs to the same unique orbit, since applying this generator repeatedly to a curve generates the entire curve.

When we are discussing symmetries of geometric objects, an important concept is that of a minimal set of representatives of the orbits of the symmetries of the entire object. Thus, by repeatedly applying the set of generators to this set, we can generate the entire object. We call this sub-object the *fundamental domain* of the object [1].

Definition. Given a diagram D on the plane with symmetry group G, a *fundamental domain* for D is a connected subset $S \subset \mathbb{R}^2$ such that for all $x \in \mathbb{R}^2$ there exists $g \in \mathsf{G}$ and a unique $y \in S$ such that $g(y) = x$; *i.e* every point in the plane has exactly one point in S that maps onto it under some isometry.

Fundamental domains are often called *fundamental regions* in the literature. For a diagram, the fundamental domain is typically chosen, where possible, to be a connected region of the underlying space, and the concept naturally generalizes to diagrams on surfaces other than the plane.

For the case of diagrams with n-fold rotational planar symmetry, a fundamental domain of the diagram can be any pie-slice formed by the section of the diagram between two straight rays from the point of rotation to infinity offset by $2\pi/n$ radians from each other. An example of a fundamental domain for this type of object is shown

[1] To be clear, the *set of generators* is the minimal subset of isometries that can be used to generate all symmetries, whereas the *fundamental domain* is the set that generates the entire diagram under the action of those generators.

in Figure 3.2 in Section 3.1, where we will explore fundamental domains for rotational symmetry more fully.

2.4.3 Cyclic Groups

Cyclic groups are in some ways the simplest groups over a set of size n.

Definition. A group G with n elements is *cyclic* if it has a single generator (but note that there may be more than one possible choice of generator), and it must have order n.

A common example of the cyclic group of order n relates to modular arithmetic: the set of integers $\{0, 1, \ldots, n-1\}$ (written $\mathbb{Z}_n$) form a cyclic group under addition modulo n, and a generator is any k, $0 < k < n-1$, that is relatively prime to n (the greatest common divisor of k and n is 1).

A cyclic group of order n is written C_n, where the group elements are implicit from context. In Table 2.2, the first and simplest class of spherical symmetry groups is notated C_n, since a diagram with an isometry $R_{\vec{c},2\pi/n}$ has a symmetry group with this single generator, and we will see many planar and spherical diagrams with this isometry, giving diagrams that realize cyclic groups of order n for many different n. For example, the Venn diagram in Figure 3.1 has 5-fold rotational symmetry, and no other symmetries, and thus realizes the cyclic group C_5.

2.4.4 Direct Product Groups

An important method of creating large groups is by taking direct products of smaller groups; this operation is often referred to as *composition* of groups.

Definition. Given two groups G and H, the *direct product group* $\mathsf{G} \times \mathsf{H}$ is the set of ordered pairs (g, h) with $g \in \mathsf{G}$, $h \in \mathsf{H}$.

The order of the group $\mathsf{G} \times \mathsf{H}$ is $|\mathsf{G}| \times |\mathsf{H}|$, as one would expect. For example, the direct product group $\mathsf{C}_4 \times \mathsf{C}_3$ is a group of order $4 \times 3 = 12$ which in fact turns out to be the cyclic group C_{12}.

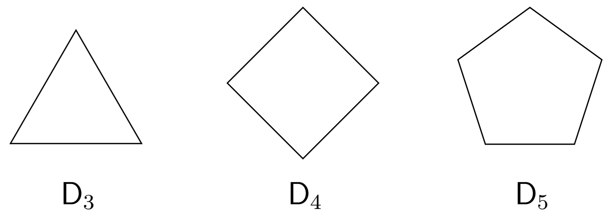


Figure 2.13: Some regular n-gons and their symmetry groups

Recall that $\mathbb{Z}_n$, the integers modulo n under addition, realize a cyclic group of order n. Then the direct product group of two cyclic groups $\mathsf{C}_i, \mathsf{C}_j$ is a set of ordered pairs (a, b) with $a \in \mathbb{Z}_i$, $b \in \mathbb{Z}_j$, which combine according to the composition rule

$$(a_1, b_1)(a_2, b_2) = (\ (a_1 + a_2) \bmod i, (b_1 + b_2) \bmod j\ )\ ,$$

which is an expression of vector addition.

We will discuss some specific direct product groups in Section 6.3.

2.4.5 Dihedral Groups

Dihedral groups are very common as they are the class of groups whose elements are symmetries of regular planar figures. For $n \in \mathbb{Z}^+$, $n \geq 3$, an *n-gon* is a regular n-sided polygon; see Figure 2.13 for some n-gons and their symmetry groups.

Definition. The set of symmetries of a regular n-gon, and its associated group, is termed the *dihedral group*, written D_n.

Let us assume we are operating on a regular n-gon centred about the origin in the Cartesian plane, in such a way that the uppermost vertex is located precisely on the y-axis (as in Figure 2.13). Any isometry of this object maps vertices to vertices, so an isometry can be exactly specified by choosing an initial labelling of the vertices and then specifying the corresponding permutation induced by the isometry. For example, label the vertices incrementally in a clockwise direction starting at 1 for the upper-most vertex, and then each isometry induces a permutation of $(1, 2, \ldots, n)$, where the permutation maps i to j exactly when the isometry maps vertex i to vertex j.

There are two generators for the group of symmetries of these objects: the rotation about the centre point by $2\pi/n$, combined with a reflection across a line of symmetry bisecting (and orthogonal to) any edge. It is simple to show that every possible isometry can be established as a finite product of one or more of these rotations and one or more reflections, and moreover that all elements of D_n have a unique representation by zero or one reflections followed by at most n rotations. Thus we have established the order of D_n as

$$|\mathsf{D}_n| \;=\; 2n.$$

In group terminology, the dihedral group D_n is isomorphic to the direct product of the two cyclic groups $\mathsf{C}_n \times \mathsf{C}_2$.

An important dihedral group is the group D_2 of order four, called the *Klein four-group*, which is isomorphic to the direct product of two cyclic groups of order two. The Klein four-group is the symmetry group of a rectangle on the plane, with the symmetry operations being two reflections across orthogonal axes and a 180° rotation. This group is one of only two groups of order four (the other being the cyclic group C_4). The Klein four-group is the smallest non-cyclic group; see Table 2.3 for other important small groups.

In Section 2.6 we will overload our notation and write D_n for a class of groups for diagrams on the sphere whose symmetry group is isomorphic to a dihedral group.

2.4.6 Symmetric Groups

The symmetric group[2] on the n-set is the group of all permutations of n symbols, and is written[3] $\mathsf{P}_n = (\mathbb{P}_n, \circ)$, where the operation $\circ$ is composition of permutations. It is easy to show that there are precisely $n!$ permutations of the n-set $\{1, 2, \ldots, n\}$ and thus

$$|\mathsf{P}_n| \;=\; n!$$

is the size of the group.

It is important to recognize that the elements of P_n are the permutations of the n-set $\{1, 2, \ldots, n\}$, not the n-set itself; for example, the elements of the symmetry group P_3 on the 3-set $\{1, 2, 3\}$ is the set of permutations

$$\mathsf{P}_3 \;=\; \{123,\ 213,\ 321,\ 312,\ 231,\ 132\}.$$

The symmetric group P_n is the largest possible group on the n-set. Its importance for us is due to a classic result known as Cayley's Theorem, which states that any finite group of order n is isomorphic to a subgroup of P_n. For example, the subgroup of size n with elements consisting of the n permutations

$$1\,2\ldots n,\;\; n\,1\,2\ldots(n-1),\;\; \ldots\;,\;\; 2\,3\ldots n\,1$$

forms a cycle of size n and is isomorphic to the cyclic group C_n.

[2] The reader is cautioned that the name of this group can cause confusion, since we are dealing with symmetries of diagrams whose symmetry groups may or may not be the symmetric group.

[3] The reader is also cautioned that the standard notation for the symmetric group is S_n; however we choose the alternate notation to avoid confusion with the polydromic symmetry S_{2n} which is extensively used throughout.

2.5 Groups on the Plane

Before discussing symmetry groups on the sphere a brief mention of plane symmetry groups is warranted; see references such as [109, 10, 133] for more details. Plane symmetry groups can be classified into three types: the *point groups*, which fix a point, the *frieze* groups, which are repetitive in one direction, and the *wallpaper* groups, which repeat in two or more directions. We have seen that the dihedral group D_n is the symmetry group of a regular n-gon, and in fact it is well-known that the dihedral groups and cyclic groups C_n are the only plane symmetry point groups (see, for example, [47, p. 457] for a proof). There are seven frieze groups, and 17 wallpaper groups.

2.6 Groups on the Sphere

In this section we give an introduction to the finite symmetry groups on the sphere. A full derivation of all the groups, while beyond the scope of this work, is a wonderful piece of algebra and geometry that is originally due to Klein [80, Chap. 5, Sec. 2]; see [26, 114] for other presentations using more modern language. The classification of symmetry groups in 3-space is an enormous topic about which entire books have been written [111], so we cannot hope to do more than present a very brief summary that vastly simplifies the various nuances involved.

Observe that the symmetries of the sphere are the isometries of Euclidean space that leave the origin fixed, and thus map a sphere centered at the origin onto itself. In Euclidean space all isometries can be classed into six types: *reflection* across a plane, *rotation* about an axis, *translation* along a vector, *screw rotation* (rotation followed by translation in a direction along the rotational axis), *rotary reflection* (rotation followed by a reflection across a plane orthogonal to the rotational axis), and *glide reflection* (translation followed by reflection across a plane containing the translation vector). An isometry of the sphere must fix its centre, so all symmetries of the sphere are types of reflection, rotation, and rotary reflection (Section 4.2 contains more precise definitions of the isometries of the sphere). The set of all symmetries of the sphere, the sphere's symmetry group, is usually termed the *orthogonal group*, written $O(3)$ (the 3 refers to the sphere in 3-dimensional space).

There is no commonly accepted notation for the finite subgroups of $O(3)$ (again, entire chapters of books have been written on notation [61], and it is difficult to read the literature without some cross-referencing chart of notation) and so we feel justified in choosing a notation similar to those in most common usage but which best suits our purposes. We use a simplified variant of Coxeter notation ([25], see also [77, Appendix 9] and [27, Table 2] for more details) and the Schönflies crystallographic notation [110] with some modifications to ensure simplicity and parallels with our group notation of earlier sections.

The subgroups of $O(3)$ naturally divide into two types. The first type is the infinite classes that have a parameter, notated n, and each class is characterized by an axis of rotation about which is an n-fold rotational symmetry. These groups include cyclic and dihedral groups, and in total they are called the *prismatic* or *frieze* groups since they refer to the symmetry of a prism (a regular n-gon, as in Figure 2.13, extruded into a 3-dimensional prism) with a pattern (or frieze) around the outside faces (created by the extrusion). The second type is the groups corresponding to the symmetry groups of the Platonic solids and their subgroups, depending on whether and how reflective planes are allowed.

We present the finite subgroups of $O(3)$ in Table 2.2, showing the name of the group which we will refer to it as, the symbolic representation, the group order, and a description and figure of an object with that symmetry group[4]. The symbols we use consist of the basic part $\mathsf{C}_n, \mathsf{D}_n, \mathsf{S}_{2n}, \mathsf{T}, \mathsf{O}$ or I, designating cyclic, dihedral, polydromic, tetrahedral, octahedral, and icosahedral rotation types, with subscripts

[4]Some figures in Table 2.2 modified from [79].

added to indicate additional mirror planes (the only exception is the polydromic symmetry, which is a prism whose frieze has a glide reflective symmetry). Generally the subscripts h and v indicate mirror planes perpendicular to and parallel to a main rotational axis taken as vertical, or the main rotational axes in the case of the fixed-size groups T, O and I. Additional subscripts are added to indicate special cases, subgroups, or those that do not fit the preceding types; see a more specific reference (for example [13]) for more details.

It is also useful to single out a few important small groups, including the trivial group of order one, which are subgroups of those in Table 2.2, as we will see diagrams exhibiting these groups; these are shown in Table 2.3.

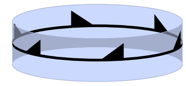
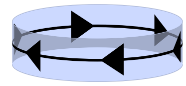
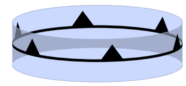
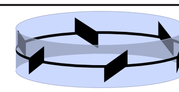
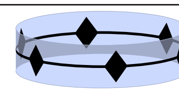
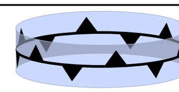
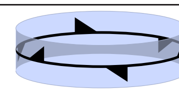
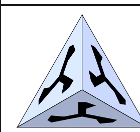
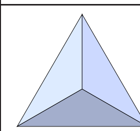
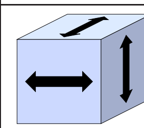
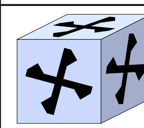
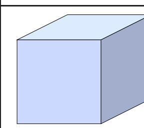
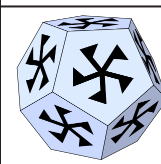
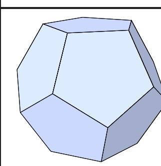
Name	Group Notation	Order	Polyhedron	Figure
polytropic (rotational)	C_n	n		
polygyros	C_{nh}	$2n$		
polyscopic (pyramidal)	C_{nv}	$2n$	regular n-gonal pyramid	
polyditropic (dihedral)	D_n	$2n$		
polydiscopic (prismatic)	D_{nh}	$4n$	regular n-sided prism	
polydigyros (antiprismatic)	D_{nd}	$4n$	regular n-sided antiprism	
polydromic	S_{2n}	$2n$		
chiral tetrahedral	T	12	tetrahedron w/o reflective	
tetrahedral	$\mathsf{T}_d \cong \mathsf{P}_4$	24	tetrahedron (with reflective)	
pyritohedral	T_h	24	pyritohedron	
chiral octahedral	$\mathsf{O} \cong \mathsf{P}_4$	24	octahedron or cube w/o reflective	
octahedral	$\mathsf{O}_h \cong \mathsf{P}_4 \times \mathsf{C}_2$	48	octahedron or cube	
chiral icosahedral	I	60	dodecahedron or icosahedron w/o reflective	
icosahedral	I_h	120	dodecahedron or icosahedron	

Table 2.2: Finite groups on the sphere, $O(3)$

Name	**Group Notation**	**Order**	**Fundamental Domain**
monotropic (no symmetry, trivial)	C_1	1	entire sphere
reflection	$\mathsf{C}_{1v} = \mathsf{C}_{1h}$	2	hemisphere of sphere
monodromic (inversion)	$\mathsf{S}_2 \cong \mathsf{P}_2$	2	hemisphere of sphere
Klein four-group	$\mathsf{C}_2 \times \mathsf{C}_2 \cong \mathsf{D}_2$	4	half of hemisphere of sphere

Table 2.3: Important small subgroups on the sphere, special cases of those in Table 2.2

Chapter 3

Symmetry in Diagrams

In this chapter we describe the basic terms and historical precedents necessary for our further study of symmetries in subsequent chapters, focussing mostly on symmetric diagrams in the plane. While the terminology and notation that other researchers has used is at times slightly at odds with our own, in this section we will use the traditional terminology and note where it differs from our own use in future sections; it will be clear from context, for example, what kind of symmetry we are referring to when we call a diagram "symmetric". Note also that the term "symmetric chain decomposition" is unrelated to a symmetric diagram, and though we will use one to create the other, it will be clear which objects are being referred to as symmetric.

3.1 Planar Rotational Symmetry

Recall the types of isometries discussed in Chapter 2, specifically rotations.

Definition. A subset $S \subseteq \mathbb{R}^2$ in the plane is *rotationally symmetric*, or *symmetric*, if it has a rotational isometry $R_{c,\theta}$ with $0 < \theta < 2\pi$.

This notion can be made more specific when applied to diagrams, as we also specify that the rotational symmetry must map curves in their entirety onto other curves.

Definition. A diagram $D = \{C_1, C_2, \ldots, C_n\}$ of n curves in the plane is *rotationally* *symmetric* if D has a rotational isometry $R_{c,2\pi/n}$, such that $R_{c,2\pi/n}(C_i) = C_{i+1 \pmod n}$ for all $1 \leq i \leq n$.

See Figure 3.1 for an example of a diagram with 5-fold rotational symmetry, with lines showing 5 equivalent sections of the diagram, showing that it is acted on by the symmetry group C_5. One fundamental domain of a figure realizing rotational symmetry group C_n in the plane is a wedge stretching outwards from the point of rotation c and forming an angle at the point of rotation of $2\pi/n$. The two sides of the wedge are defined by two copies of an arc starting from c and going to infinity, one copy is separated from the other by the rotation of $2\pi/n$, and the arcs are such that they do not intersect each other. One arc defines the closed boundary and the other the open boundary, one on either side. Figure 3.2 shows a fundamental domain for the example diagram from Figure 3.1. The region between the two arcs is referred to as a *sector*, and informally several authors have referred to it as a "pie-slice" for obvious reasons.

It is useful to recall, from the definition of a fundamental domain, that any fundamental domain for a diagram realizing the symmetry group C_n under the rotation isometry $R_{c,\theta}$ can be chosen as the fundamental domain for the diagram, an idea which has been important in previous work in searching for symmetric Venn diagrams ([19]). The two arcs, one a copy of the other rotated by θ around c, that define the boundaries of the fundamental domain, do not necessarily have to be linear; and in fact the fundamental domain could be defined by more than one pair of arcs, given more than one "wedge", which are disjoint (except at the central point c), though in all of our work the fundamental domains we use are connected.

The diagram in Figure 3.3(a) is rotationally symmetric: if we rotate the curves $2\pi/4 = \pi/2$ clockwise about the centre point and permute the curves by the circular permutation (*red* $\rightarrow$ *blue* $\rightarrow$ *green* $\rightarrow$ *red*), we obtain the same diagram. The diagram in Figure 3.3(b) is not rotationally symmetric, despite its having left-to-

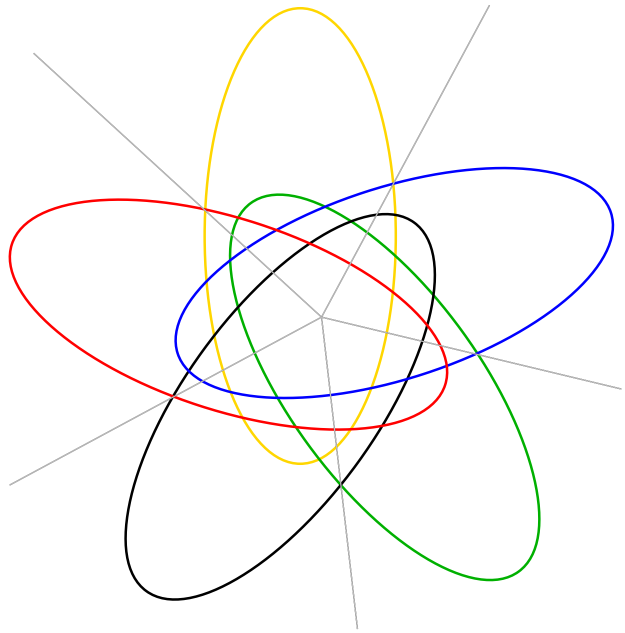


Figure 3.1: A diagram with planar rotational symmetry, realizing C_5

right mirror-image symmetry.

An obvious necessary condition for a diagram to be rotationally symmetric is for all curves to be congruent. Implicit in the definition is that the curves $\{C_1, C_2, \ldots, C_n\}$ can be labelled 1 to n so that $R_{c,2\pi i/n}(C_j) = C_{(j+i \pmod n)}$, and thus the permutation induced by this operation is a single cycle of length n; see [107] also. In general, the operation of a rotation is equivalent to relabelling the curves by a permutation of the curve labels. Futhermore we can label the curves so that the circular permutation of the curves induced by rotation is $(1\ 2\ 3\ \ldots\ n)$.

Recall the discussion in Section 2.1 that a k-region in a diagram is a region enclosed by exactly k curves. In a symmetric diagram, any k-region, after applying the symmetry operation $R^i_{c,2\pi/n}$ for $i \geq 1$, must map onto a k-region, since a region r, if it is enclosed by the curves $\{C_{x_1}, C_{x_2}, \ldots, C_{x_k}\} \subseteq \mathcal{C}$ with $1 \leq x_i \leq n$, after applying $R^i_{c,2\pi/n}$ will be enclosed exactly by $\{C_{x_1+i \pmod n}, C_{x_2+i \pmod n}, \ldots, C_{x_k+i \pmod n}\}$.

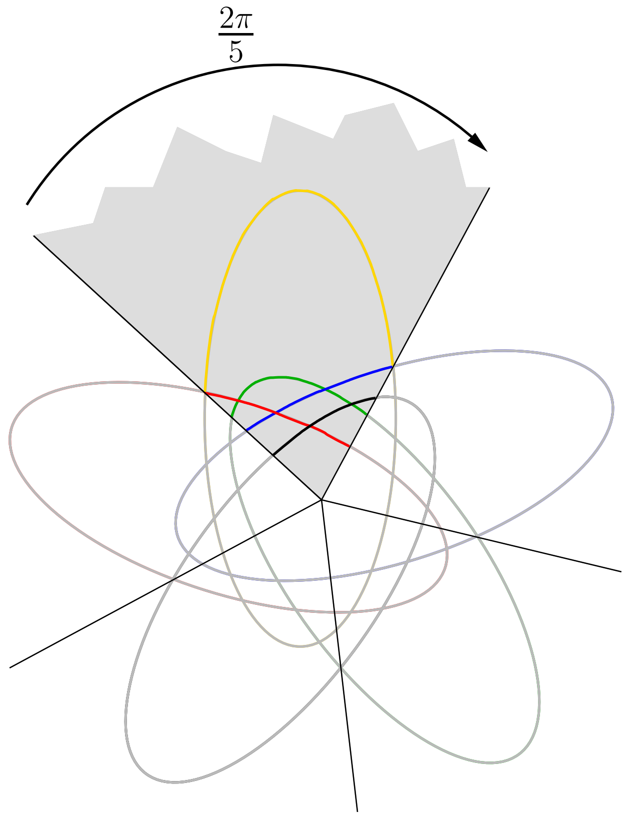
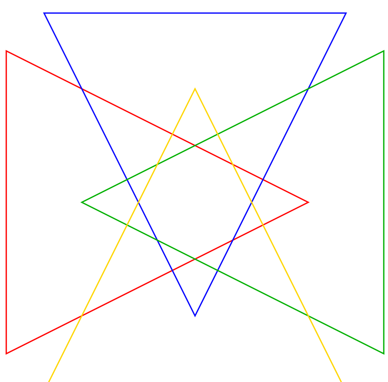
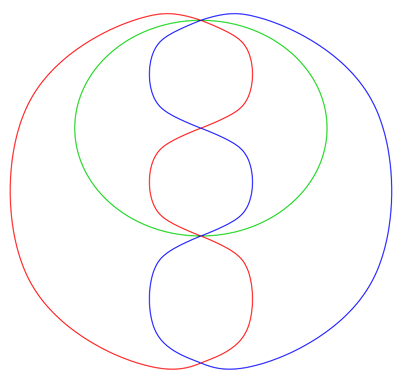


Figure 3.2: Fundamental domain for diagram in Figure 3.1



(a) A symmetric 4-curve (non-Venn) diagram

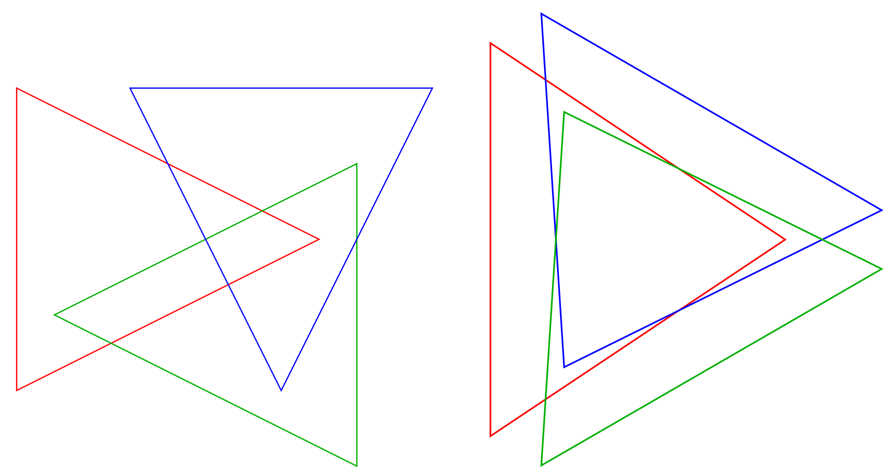


(b) A non-symmetric Venn diagram

Figure 3.3: A symmetric diagram and non-symmetric diagram

In a symmetric Venn diagram or Euler diagram, there is exactly one 0-region and one n-region, namely the unbounded external region (exterior to all curves) and the internal region (interior to all curves), these regions must map onto themselves under any rotation, and thus the point of rotation must be contained within one of these two regions. In a symmetric Venn diagram or independent family it is easy to see (though more difficult to prove, see [127] and the discussion in Section 3.2) that the point of rotation must be contained within the internal face.

As a final aside, to add to the terminological confusion, some early authors called diagrams with congruent curves (but not necessarily rotational symmetry) "symmetric". Grünbaum clears up some of this confusion in [58] by using the term *nice* to refer to a diagram that is not necessarily symmetric by rotation but is composed of congruent curves; see Figure 3.4 for an example.



(a) A nice but not symmetric diagram

(b) A symmetric and nice diagram

Figure 3.4: A nice diagram versus a symmetric diagram, both formed of 3 congruent triangles

3.1.1 Symmetric Duals

The dual of a symmetric diagram also can be drawn to have the isometry $R_{c,\theta}$ (ignoring edge labels) if we ignore the node labelled with the empty set, corresponding to the external unbounded face of the diagram. This is apparent in Figure 2.8(b) on page 24: if we remove the upper-most node and the edges incident to it, the rest of the dual diagram is symmetric. In the dual of a Venn or Euler diagram the point of rotation is about the centre node which corresponds to the region enclosed by all curves.

One must be careful when considering how a fundamental domain of a symmetric diagram translates into a fundamental domain in a dual. Consider what happens if the borders of a fundamental domain are chosen to pass through a region r in the diagram. Let r' be the part of the region contained inside the fundamental domain, and r'' be the part outside of the fundamental domain; see Figure 3.5 for an example, showing the dual of the fundamental domain from Figure 3.2. Then the dual of the section of the diagram contained within that fundamental domain will contain a node corresponding to r', and a node $R_{c,\theta}(r'')$ corresponding to an image of r'' under one application of the isometry or its inverse ($R_{c,\theta}$ or $R_{c,\theta}^{-1}$). Applying the symmetry operation n times to generate the entire diagram would then produce $2n$ images of these two nodes; however, both of these nodes are dual to the single region r, which has an orbit of size n in the final diagram, and the regions in this orbit will have n nodes in the dual.

These issues have been treated by most authors [54, 64, 65] by the following techniques:

- First, one can consider the fundamental domain of the dual of a diagram to not include nodes corresponding to the internal and outer faces. This means that once the symmetry operation is applied to the fundamental domain to create the entire dual graph, two extra vertices have to be added in to create the full dual. There is no ambiguity as to the location of these extra vertices since the external face in the diagram will go to infinity and the internal face contains the point of rotation (and the dual node corresponding to the internal face is conventionally located at the point of rotation).

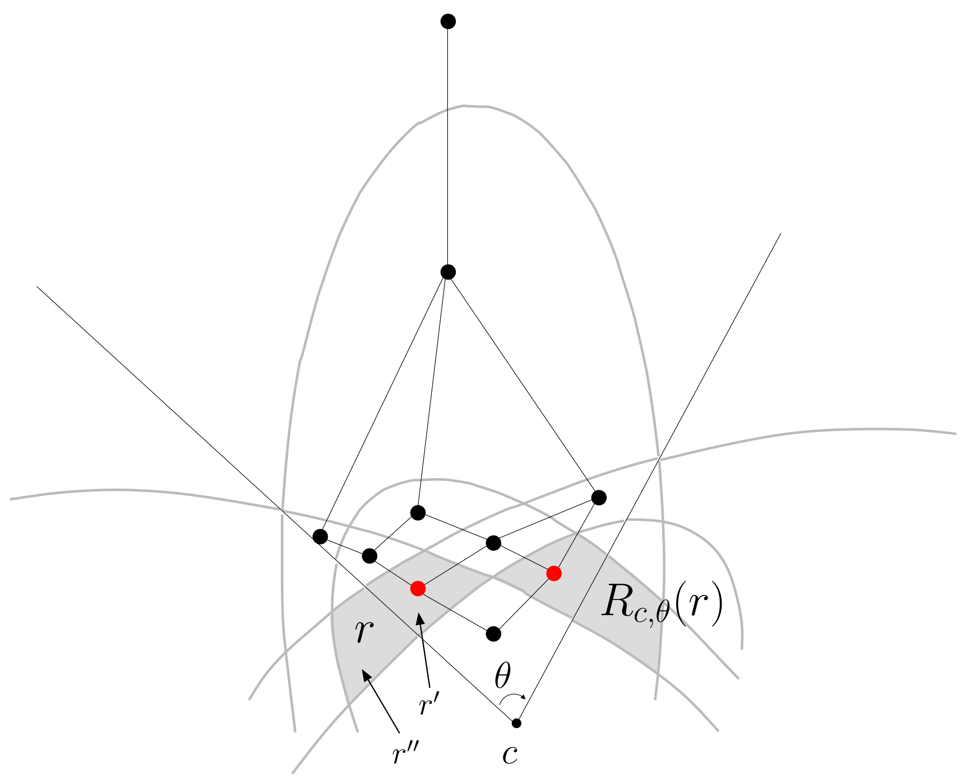


Figure 3.5: A fundamental domain that spans a region creates multiple nodes for that region in its dual

- One can choose the boundary of a fundamental domain of a diagram to be coincident with edges and vertices in the diagram, thus ensuring that the dual of the fundamental domain will have exactly one node for each region represented. Moreover, if all of the curves of the diagram pass through a vertex, then the fundamental domain can be chosen to be the sector of the diagram between two successive copies of this vertex (this is the approach used by the two well-known constructions for large symmetric Venn diagrams, the GKS construction (discussed later) [54] and Hamburger's work [64]).

Since any choice of two arcs that are non-intersecting and congruent under a rotation about the point of rotation can be used to define a fundamental domain, as discussed earlier, the borders of the fundamental domain can be altered so that they do not pass through any region in the diagram, except the internal and external regions.

3.1.2 Necklaces and Dual Fundamental Domains

Since the rotation operation induces a cyclic permutation in the curves, the labels of regions also are permuted by this cyclic permutation when the region undergoes the rotation operation. An important set of equivalence classes of the power set B_n is defined by this operation of rotation on strings. Here we define and describe these classes, called necklaces, enough for our purposes in describing the GKS construction. Later, in Section 7.1, we will more fully describe necklaces in terms of shift registers for other types of symmetric diagrams.

For $x = x_1 x_2 \ldots x_n \in B_n$, let $PCR(x)$ be the "rotation" function on x defined by $PCR(x) = x_2 x_3 \ldots x_n x_1$. The PCR notation is due to the description of this function as the *pure cycling register* and analysing it in terms of shift register sequences, in [51]. Rotating x multiple times is notated $PCR^i(x) = PCR(PCR^{i-1}(x))$, where $PCR^1(x) = PCR(x)$. Now define the relation $\stackrel{PCR}{\sim}$ on B_n by $x \stackrel{PCR}{\sim} y$ iff $y = PCR^i(x)$

for some $i \geq 1$. Then $\overset{PCR}{\sim}$ is an equivalence relation on B_n, which we call the *necklace* relation, and the equivalence classes are called *necklaces*, because each string can be thought of as a circular sequence of bits on a rotatable chain.

Let N_n be the set of necklaces of B_n. Define the *necklace poset* $\mathcal{N}_n$ by $\mathcal{N}_n = (N_n, \preceq)$ where the ordering $\preceq$ is defined for two necklaces ν_1, ν_2 by $\nu_1 \preceq \nu_2$ iff there exists $x \in \nu_1$ and $y \in \nu_2$ such that $x \prec y$ in B_n, *i.e*, x and y differ in some non-empty set of bit positions where x has 0s and y has 1s.

Thus, since the equivalence classes of strings under cyclic permutations are necklaces, the set of node labels in a fundamental domain for the dual of a diagram will be exactly a set of necklace representatives. Under the above conventions, this set of necklace representatives will not include 0^n or 1^n, since these two strings each belong to equivalence classes of size one, one each for the internal and external faces. At the risk of inducing unnecessary confusion, we will assume whenever we are discussing planar symmetric diagrams, unless otherwise indicated, that the necklace poset $\mathcal{N}_n$ does not include 0^n and 1^n.

Necklaces are a well-studied combinatorial object, with well-known fast generation algorithms and other related operations known for several decades [44, 45, 105]. Their applicability to generating Venn diagrams by generating the dual, however, has only been utilized recently, as discussed in the following sections. We further discuss necklaces in Section 7.1, where some variants will be introduced.

3.2 History of Symmetry in Venn Diagrams

Now that we have established the necessary terminology and basic concepts, we review the history of research in symmetry as it applies to diagrams on the plane.

Work on symmetry of diagrams has mostly focussed on Venn diagrams and near-Venn diagrams (with close to 2^n faces), and arguably started with the key results of Henderson's brief 1950 paper [72], in which he showed the following theorem.

Theorem 1 ([72])**.** *For an n-Venn diagram to be symmetric, n must be prime.*

For an understanding of why this is true, consider the $\binom{n}{k}$ connected subsets of weight k, $0 < k < n$, and note that the centre of rotation must be contained with the internal face, as the external (empty) face and internal face (contained by all curves) must map onto themselves by the symmetry operation. Then the k-subsets must be distributed evenly about the centre of rotation for any rotation by $2i\pi/n$, $0 \leq i \leq n$ to map them onto themselves. Thus they must be divided into n orbits, each containing a whole number of regions, so n must divide $\binom{n}{k}$ for $0 < k < n$. This implies n must be prime, by a theorem of Leibnitz [71].

This somewhat intuitive argument has some subtle problems that have recently been corrected with a proof that provides a more solid formal footing by Wagon and Webb [127]. Their argument avoids the binomial coefficients altogether: assuming a symmetric diagram exists, a contradiction is derived in the situation when $n = \alpha \times \beta$ is composite by some clever group theory applied to the subgroups of rotation of order α and β. In [127] the authors also rigorously prove that the centre of rotation in a symmetric Venn diagram or independent family must be contained in the centre, internal face.

The study of symmetric diagrams was revived by Grünbaum with a series of papers, starting with his award-winning paper from 1975 [56]. In this work he proves some fundamental bounds relating to symmetric independent families and provides an early construction for an n-Venn diagram that exhibits some nice symmetries; it is useful to note that his construction is isomorphic to the construction of Edwards [37, 38] 14 years later, though Grünbaum's uses polygons and rapidly gives very small regions as n increases whereas Edwards' uses smooth curves. Grünbaum was also first to find the unique simple symmetric Venn diagram of 5 curves (this diagram and other non-simple diagrams were found independently by later authors [113, 134]), shown in Figure 3.1, which corrects a misstatement of Venn in [125], in which he appeared to claim that a 5-Venn diagram of ellipses did not exist. The paper [56] also exhibits some symmetric simple and non-simple independent families and Venn diagrams of five and seven curves; at the time it was the most comprehensive and in-depth work on Venn diagrams since Venn [125].

3.2.1 Constructions and Symmetry

Numerous 20th-century authors have provided constructions, for any n, of Venn diagrams of n curves [2, 7, 9, 90, 92, 100]. Most of these constructions exhibit some planar symmetries, if the first one or two curves (which usually divide the plane vertically and horizontally) are ignored. The construction of Poythress and Sun [96] displays the type of rotational symmetry that we will later call *total symmetry* of order four, but no reflective isometries. Besides the construction due to Venn, perhaps the most well-known (due in part to the book [34]) is that of Edwards [37, 38]. Edwards' construction uses smooth curves and, as Edwards notes in passing in [34], embedding the construction on the sphere gives more nice symmetries than on the plane (though his construction on the sphere was only explored up to $n = 4$ curves). We explore this notion further in Section 8.1.

The construction of Fisher et al. [41] is isomorphic to Edwards' construction, though it uses zigzags instead of smooth curves. Glassner [50] discusses constructions isomorphic to Edwards', with many nice variations.

In [59], Grünbaum explored various symmetric diagrams formed from geometric shapes, including five ellipses, five equilateral triangles, five rectangles, seven non-convex pentagons, and made the following important conjecture, the consequences of which drove much of the subsequent research in symmetry in diagrams.

Conjecture 1 ([59])**.** *For every positive prime n there exist planar symmetric n-Venn diagrams.*

This conjecture was not answered in the affirmative until 2004 by Griggs, Killian and Savage [54], as we discuss in Section 3.3.2. He also made the following, related, conjecture.

Conjecture 2 ([59])**.** *For every positive integer n there exist Venn diagrams formed of n congruent curves.*

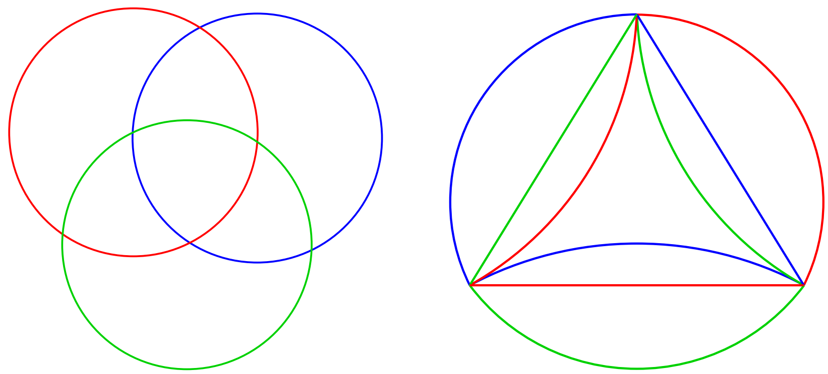
This conjecture remains unanswered, except for a construction using a specific generalization of our diagrams in which curves are allowed to have infinite intersections [106]; we also use this generalization in Section 8.3.2.

Many different methods have been used to exploit symmetry in searching for interesting Venn diagrams and independent families and drawing them nicely on the plane; for surveys of general techniques see [19, 60, 108].

The 1990s saw a flurry of activity in research surrounding diagrams, symmetric and otherwise. First, several researchers produced a definitive record of symmetry in simple Venn diagrams of seven curves. The first results, found independently by several researchers, were constructed by hand [33, 35, 59], and later computer searches exhaustively generated various types [106, 131]. Ruskey [106] found 33 simple non-monotone symmetric 7-Venn diagrams, and it is suspected that there are no more. It was also shown in [106] and verified by Cao [14] that there are exactly 23 simple symmetric 7-Venn diagrams that are monotone.

Also in the 1990s, Peter Hamburger and his coauthors published a series of papers [16, 17, 18, 67, 68] establishing several graph-theoretic properties of Venn diagrams. As part of this work they counted the total number of simple 5-curve Venn diagrams with different properties, and in so doing established that there are is only one simple symmetric five-curve Venn diagram. They also established the total number of three-curve Venn diagrams and found that there are only two symmetric three-curve diagrams, those shown in Figure 3.6.

Hamburger also found, by hand, the first symmetric 11-Venn diagram [64], using a heuristic technique. By modifying the diagram in various ways, he also created



(a) The simple symmetric 3-Venn diagram

(b) The only other symmetric 3-Venn diagram

Figure 3.6: The two unique symmetric three-curve Venn diagrams

variants with different vertex counts [62, 63, 65, 69]. His work also uses some of the ideas discussed in Section 3.1; he uses the terms *generator* to refer to a set of necklace representatives used to build the fundamental domain for the dual graph, which itself he terms a *doodle*.

3.2.2 Polar Symmetry

Grünbaum was the first researcher to discuss a second type of symmetry, related to stereographic projections (defined in the following chapter). This symmetry has proved fruitful in research in finding symmetric Venn diagrams, for practical and aesthetic reasons.

Definition. Let D be a planar symmetric diagram. Consider the operation of

- stereographically projecting D onto a sufficiently-large sphere with the sphere's south pole tangent to the plane at the point of rotation, and then
- stereographically projecting D onto the (parallel) plane tangent to the opposite (north pole),
- taking the mirror reflection of the resulting diagram across any line to give a diagram D'.

If D' is isomorphic to D, then D has *polar symmetry*.

Intuitively the projection onto the plane and back onto the parallel plane tangent to the opposite pole is equivalent to "turning D inside out" on the plane; in geometry, this operation is called *circle inversion* [129], in which the diagram is reflected across a circle centred at the point of rotation. The terms *self-complementary* [60] and *spherical* [59] symmetry have also been used to refer to the polar symmetry concept.

More properly, the term polar symmetric also applies to diagrams that are isomorphic by just considering the circular inversion and not its mirror reflection; however there are very few diagrams, and those that we know of have very few curves, that are polar without applying the mirror reflection, so in this work we use the term *polar symmetry* to apply exclusively to the symmetry attained by applying the circle inversion to the diagram, as above, and then applying a reflection to achieve D'. If D' is isomorphic to D, then any translation and rotation of D' is as well, which means that any line of reflection can be chosen to take the mirror reflection, since the results of two different reflections only differ by translation and rotation. Note that the polar symmetry operation is not an isometry (though, as we will see later, on the sphere it translates to an isometry). Note also that the curves in the diagram are all congruent due to the rotational symmetry. The polar symmetry also applies to the dual graph of the diagram, and since the dual graph can map onto itself under the polar symmetry operation and each curve maps onto some other curve under the isometry, each curve can map onto any curve (including itself) under the symmetry since they are all congruent, and continue to be after the polar symmetry operation.

First discussed by Grünbaum in [58], polar symmetry was used to guide many of the early searches for rotationally symmetric Venn diagrams by Edwards [35] and Ruskey [19, 106], since restricting a search for symmetric diagrams to those with polar symmetry can dramatically reduce the computation involved since only half of the diagram needs to be generated (more precisely, the size of the fundamental domain involved is halved; a further discussion of computation issues can be found in [19, 131]). Of the 23 simple symmetric monotone 7-Venn diagrams, six are polar symmetric, which were the first to be discovered. There are not known to be any non-monotone polar symmetric 7-Venn diagrams. Some earlier diagrams were further explored in regards to polar symmetry [59]; the two rotationally symmetric 3-Venn diagrams, shown in Figure 3.6, are both polar symmetric, as is the simple 5-Venn diagram of ellipses in Figure 3.1.

We will discuss polar symmetry further in the context of curve-preserving symmetries on the sphere in Chapter 4 since the notion of polar symmetry translates to a nice spherical symmetry operation.

3.3 Constructing Venn Diagrams from Chain Decompositions

In this section we introduce the important results of Griggs, Killian, and Savage [54] of 2004 in which they proved, by construction, that planar symmetric Venn diagrams exist for any prime n. We use the same basic technique in Chapter 5 to create various diagrams from chain decompositions, and their work provides the theoretical underpinning for why the construction works, and similar techniques are used in Chapter 7. Hence, this section consists primarily of an overview of [54], with emphasis on the results we will use later; we refer interested readers to the original paper [54] or [108] to fill in the details.

We refer to these constructions, both for (monotone) non-symmetric and symmetric Venn diagrams, as the "GKS construction" throughout the remainder of this work.

3.3.1 Monotone Venn Diagrams from Chain Decompositions

Before discussing the symmetric Venn diagram construction, the authors of [54] prove an important preliminary result, the mechanics of which we will use in Chapter 5 to build Venn diagrams on the sphere with different isometries. Some of the definitions in this section will be modified later with terminology suited to our later purposes.

Recall the poset $\mathcal{B}_n$ called the boolean lattice, whose elements are B_n, the set of n-bit strings (corresponding with subsets of the n-set). The first contribution of [54] is to show how the symmetric chain decomposition of $\mathcal{B}_n$ we discussed in Section 2.3.3, due to de Bruijn, et al. [28], gives a straightforward construction for monotone Venn diagrams for n curves.

Let $\mathcal{C}$ be a symmetric chain decomposition of $\mathcal{B}_n$ and for chain $C \in \mathcal{C}$, let $starter(C)$ be the first element of C and let $terminator(C)$ be the last element of C. Call the longest chains in C the *root chains*.

Definition ([54, p. 6])**.** A chain decomposition $\mathcal{C}$ has the *chain cover property* if whenever $C \in \mathcal{C}$ and C is not a root chain, then there exists a chain $\pi(C) \in \mathcal{C}$ such that

$starter(C)$ is covered by a starter element $\pi_s(C)$ of $\pi(C)$, and

$terminator(C)$ covers an element $\pi_t(C)$ of $\pi(C)$,

where π is called the *chain cover mapping*.

Given the string $x \in B_n$, regard the 0s as left parentheses and 1s as right parentheses, and match parentheses in the usual way. That is, as x is scanned from left to right, when a 0 is encountered, it becomes an unmatched 0, and when a 1 is encountered, it is matched to the rightmost unmatched 0 to its left, if any, otherwise it becomes an unmatched 1. Recall that the chain starters are those strings with no unmatched 1s, and the terminators are those with no unmatched 0s. Let $U_0(x)$ and $U_1(x)$ represent the sets of indices of unmatched 0s and unmatched 1s, respectively, in the string x. Let $S(x) = \{i : x_i = 1, 1 \leq i \leq n\}$ be the set of indices of 1s in x.

Let $\mathcal{C}$ be a symmetric chain decomposition of $\mathcal{B}_n$. Recall the Greene-Kleitman formulation of the de Bruijn chain decomposition, discussed in Section 2.3.3. We can define the chains by their initial strings, which are the strings with no unmatched 1s:

$$\mathcal{C} = \{C_x \mid x \in B_n, U_1(x) = \{\}\} \ .$$

Given x with no unmatched 1s, the entire chain can be generated by starting with x and moving left-to-right, changing unmatched 0s to 1s. The final string in the chain will have all unmatched 1s in place of the initial unmatched 0s.

Lemma 1 ([54, p. 18])**.** *The Greene-Kleitman symmetric chain decomposition of $\mathcal{B}_n$ has the chain cover property, with chain cover mapping π, defined for $C_x \in \mathcal{C}$ with $S(x) \neq 0$ by*

$$\pi(C_x) = (C_y) \ ,$$

where $S(y) = S(x) - \{max(S(x))\}$.

The following two lemmas establish that, given a symmetric chain decomposition with a chain cover mapping, we can use this to embed in the plane all of the chains and use them as the dual of a Venn diagram.

Definition. Given a chain decomposition $\mathcal{C}$ with the chain cover property, the *chain cover graph*, $G(\mathcal{C}, \pi)$ is a graph whose vertices are the elements of the poset that $\mathcal{C}$ covers, and whose edges consist of the covering relations in the chains in $\mathcal{C}$ together with the *cover edges*, for each chain C, from $starter(C)$ to $\pi_s(C)$ and from $terminator(C)$ to $\pi_t(C)$.

This chain cover graph can now be embedded on the plane, according to the following lemma.

Lemma 2 ([54, p. 7])**.** *Let $\mathcal{C}$ be a symmetric chain decomposition with the chain cover property for poset $\mathcal{X} = (X, P)$, and let π be a chain cover mapping for $\mathcal{C}$. The chain cover graph $G(\mathcal{C}, \pi)$ has a planar embedding $P(\mathcal{C}, \pi)$.*

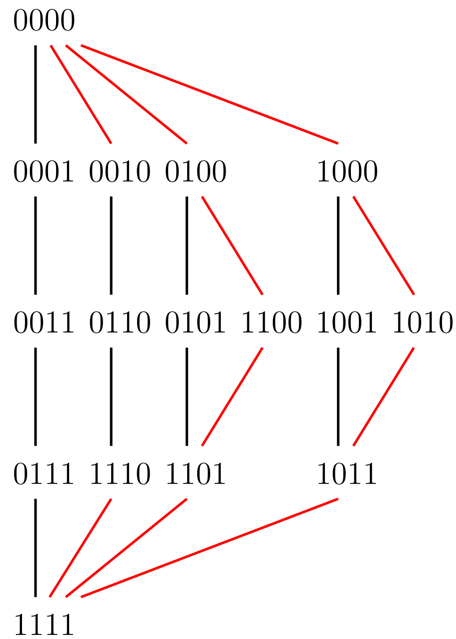


Figure 3.7: The planar embedding from [54] of the symmetric chain decomposition of $\mathcal{B}_4$, with chain edges shown in black and cover edges between chains shown in red

The proof in [54, p. 8] specifies the coordinates of the layout of each vertex and specifies how each edge is drawn, as a straight line between its endpoints. Chains are laid out as vertical paths, with all vertices from the same level in the poset lined up horizontally. A chain that is covered by another chain is embedded as close to it as possible, and multiple chains covered by the same parent are ordered shortest chain to longest. If the chains are thought of as dangling from a tree and attached by their topmost cover edges, the root of the tree is the root chain and the chains are shorter as we move further down the tree; by laying out this tree in a planar fashion, the cover edges at the bottom of each chain exactly mirrors the top, and so the entire tree structure is duplicated on the bottom of the chains, except inverted, and will also be planar.

See Figure 3.7 for an example of the planar embedding of the chain cover graph for the Greene-Kleitman symmetric chain decomposition.

Lemma 3 ([54, p. 8])**.** *Let $\mathcal{C}$ be a symmetric chain decomposition with the chain* *cover property for $\mathcal{B}_n$, let π be a chain cover mapping for $\mathcal{C}$, and let $P(\mathcal{C}, \pi)$ be the planar embedding of $G(\mathcal{C}, \pi)$ described in the proof of Lemma 2. Then the geometric dual of $P(\mathcal{C}, \pi)$, $P^*(\mathcal{C}, \pi)$, is a monotone Venn diagram of n curves with the minimum number of vertices.*

Again, the proof of Lemma 3 in [54, p. 8-11] is technical but fairly intuitive. Each vertex of $P(\mathcal{C}, \pi)$ corresponds to a region in the Venn diagram, whose label is the label $x \in B_n$ of that vertex. Edges are coloured so that curve C_i is made up of the edges crossing cover relations $a \prec b$ such that a and b differ exactly in the bit indexed i, so the edge (a, b) is what is referred to as an *i-edge*. The mechanics of the proof include: the geometry of the edges of C_i (where each edge starts, ends, and crosses the edge between a and b), showing that the edges of C_i form a simple (non-self-intersecting) cycle, and that each region defined by the interiors and exteriors of C_i occur exactly once corresponding to the vertices of $P(\mathcal{G}, \pi)$. To show that C_i forms a simple cycle it suffices to show that the i-edges form a *bond*, that is, a minimal set of edges who removal disconnects the graph.

Figure 3.8 shows the final steps in the GKS construction by adding the planar dual to the chain cover graph in Figure 3.7.

The previous lemmas, combined with the construction of the Greene-Kleitman symmetric chain decomposition for any $n \geq 1$, combine to give the following result.

Lemma 4 ([54, p. 19])**.** *For any $n \geq 1$, a monotone Venn diagram with minimum number of vertices can be obtained as the dual of any planar embedding of the planar graph $G(\mathcal{C}, \pi)$, where $\mathcal{C}$ is the Greene-Kleitman symmetric chain decomposition of $\mathcal{B}_n$ and π is the chain cover mapping for $\mathcal{C}$ defined in Lemma 1.*

The resulting diagrams have a monochrome symmetry, which we discuss in Section 8.3.

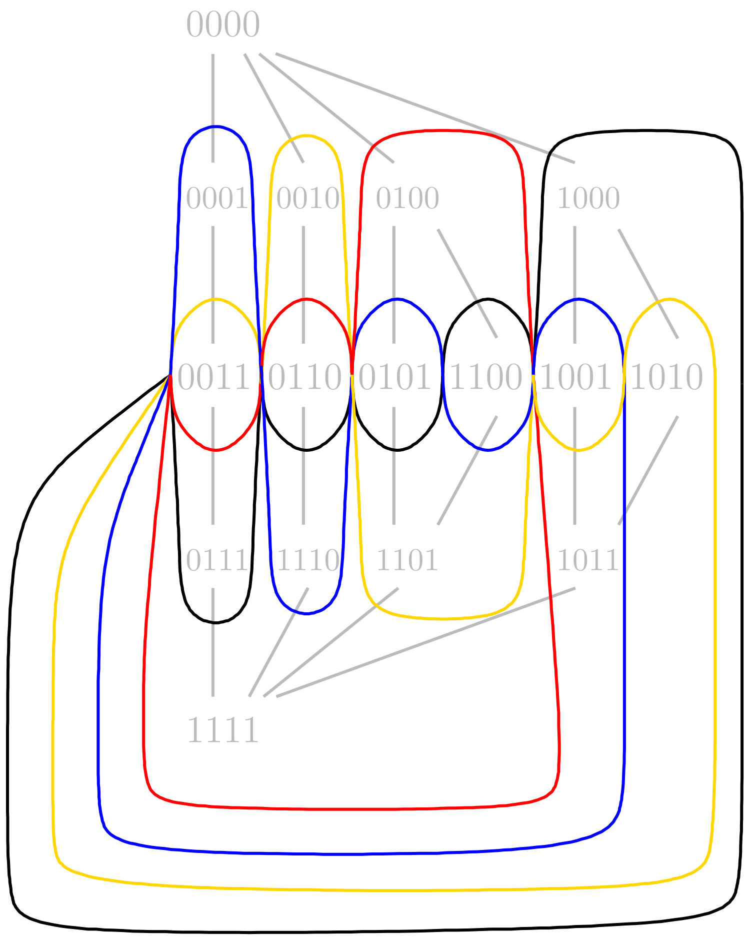


Figure 3.8: The monotone 4-Venn diagram obtained by taking the planar dual of the graph in Figure 3.7 and colouring each edge according to the index of the bit that changes over the cover relation edge crossed by that edge

3.3.2 Planar Symmetric Venn Diagrams for Any Prime n

The ideas from the previous section can be extended to build planar symmetric Venn diagrams under certain special conditions. Since the diagrams to be constructed have rotational symmetry group C_n, the fundamental domain for the entire diagram encompasses $1/n$th of the entire figure. As per the discussion in Section 3.1, we can construct this fundamental domain by constructing the dual, which will consist of nodes labelled by a set of representatives from $\mathcal{N}_n$.

To find a set of representatives of $\mathcal{N}_n$ that will build the dual, virtually the only productive technique that has been used, besides ad hoc methods, has been to search for a chain decomposition of the poset $\mathcal{N}_n$ that can then be drawn in the plane and joined together with extra edges to form a 2-connected graph; these extra edges are the cover edges from the previous section.

To summarize, the general technique for constructing a symmetric n-Venn diagram is a process as follows:

1. given n, find a symmetric chain decomposition for $\mathcal{N}_n$,
2. lay out the symmetric chain decomposition for $\mathcal{N}_n$ in the plane as a series of embedded paths,
3. connect the paths with cover edges to create a 2-connected plane graph forming the dual of the fundamental domain (a sector forming $2\pi/n$ of the final graph),
4. rotate the pie-slice by applying $R_{c,2\pi/n}$ about the centre point of rotation $n-1$ times, permuting the point labels by the cyclic permutation induced by the curve labels for each new pie-slice,
5. take the dual graph of the full structure, labelling the regions (and, implicitly, the curves) by the labels of the vertices in the dual, to recover the full symmetric Venn diagram.

The difficult part in past research has been to find the right graph for Steps 1 and 2, and various diagrams have been found by building the correct fundamental domain of the dual graph via ad hoc methods ([59, 64, 65]). The heart of the result in [54] is finding the general construction for the symmetric chain decomposition of $\mathcal{N}_n$; the rest is simply rote procedure, and in fact can be automated.

This procedure is encapsulated in the following lemma:

Lemma 5 ([54, p. 11])**.** *Let n be prime. If there exists a set R_n of necklace representatives for B_n such that the subposet $\mathcal{R}_n = (R_n, \leq)$ of $\mathcal{B}_n$ has a symmetric chain* *decomposition with the chain cover property, then a symmetric Venn diagram of n curves can be constructed.*

The proof is by construction, given the chain decomposition and chain cover edges. After fixing the point of rotation in the plane, the dual diagram is constructed by embedding the chain decomposition and chain cover edges in one sector forming the dual fundamental domain of the entire diagram. The dual is then rotated through $2\pi i/n$ radians for each $1 \leq i \leq n-1$, and at each rotation the vertices are relabelled so that a vertex labelled x becomes $PCR^i(x)$ in the ith rotation. The vertex 1^n is located at the centre of rotation (and coincides for each rotation) and 0^n is at infinity. Constructing the Venn diagram from this complete dual graph then proceeds as in the proof of Lemma 3.

The resulting diagram will be monotone, since each vertex in the dual whose label has weight k is adjacent to a vertex of weight $k-1$ and $k+1$, due to the dual of the fundamental domain being constructed from a symmetric chain decomposition. It also has the minimum number of vertices for monotone Venn diagrams, $\binom{n}{\lfloor n/2 \rfloor}$ (as in [11]).

Finally we show how the symmetric chain decomposition of $\mathcal{N}_n$ is found, by using the associated sequences called *block codes*.

Definition ([54, p. 20])**.** Given $x \in B_n$, the *block code* of x, written $\beta(x)$, is a sequence over $\{2, \ldots, n, \infty\}$ defined as follows: If $x = 1^{\alpha_1}0^{\omega_1}1^{\alpha_2}0^{\omega_2}\ldots 1^{\alpha_k}0^{\omega_k}$ with $\alpha_i > 0, \omega_i > 0, 1 \leq i \leq k$, then

$$\beta(x) = (\alpha_1 + \omega_1, \alpha_2 + \omega_2, \ldots, \alpha_k + \omega_k)$$

otherwise $\beta(x) = \infty$.

The following lemma establishes the uniqueness of block codes exactly when n is prime. A block code is *finite* if it is not ∞.

Lemma 6 ([54, p. 21]). *No two strings of B_n in the same necklace equivalence class have the same finite block code if and only if n is prime.*

This lemma enables us to define the following subposet of $\mathcal{B}_n$.

Definition ([54, p. 21]). Let $\rho(x) = \{y : y \overset{PCR}{\sim} x$ and $\beta(y) \leq \beta(z)\ \forall z \overset{PCR}{\sim} x\}$, *i.e* $\rho(x)$ is the string with minimum block code equivalent by rotation to x. Then $R_n = \{\rho(x) | x \in B_n\}$ is the set of such strings for all classes, and $\mathcal{R}_n = (R_n, \leq)$ is the subposet of $\mathcal{B}_n$ induced by R_n, called the *necklace representative poset*.

The usefulness of the block code structure is thus that it provides a method of building a symmetric chain decomposition of the necklace representative poset.

Theorem 2 ([54, p. 23]). *If n is prime, $\mathcal{R}_n$ has a symmetric chain decomposition with the chain cover property.*

The chain decomposition again uses the Greene-Kleitman chain decomposition and the parentheses matching algorithm used in the previous section to create a chain decomposition for $\mathcal{B}_n$. For $\mathcal{R}_n$, the chain starters for the chain decomposition are the strings with exactly one unmatched 1 in the first position, and the chains are built, as before, by successively changing unmatched 0s to 1s moving left-to-right in the string.

From Lemma 5 and Theorem 2 come the following result:

Theorem 3 ([54, p. 25]). *n-Venn diagrams with planar rotational symmetry exist for all prime n.*

Chapter 4

A Framework for Symmetric Spherical Diagrams

In this chapter we begin to explore representation, notation, and formalization issues surrounding diagrams and their isometries on the sphere. A more complete discussion of the spherical isometry operations provides the basis for defining and exploring basic issues of the different types of symmetries offered by coloured diagrams.

4.1 Representation and Projections

In this section we consider some issues of drawing spherical diagrams in a planar fashion, for the purpose of easily visualizing their spherical representation. We will often be considering diagrams on the sphere with a single axis of rotation, akin to the Earth, and so we refer to the points at which an axis of rotation meets the surface of a sphere as *poles*; similarly, the circumference of the sphere intersected by a plane through the centre of the sphere and orthogonal to an axis of rotation is referred to as the *equator*, as in Figure 4.1. Circumferences of the sphere perpendicular to the equator and passing through the poles are referred to as circles of *longitude*. Circles on the sphere given by the intersection of the sphere with planes orthogonal to the axis of rotation are referred to as circles of *latitude*; the equator is the line of latitude of maximum radius. Using a standard Cartesian coordinate system the z (vertical)

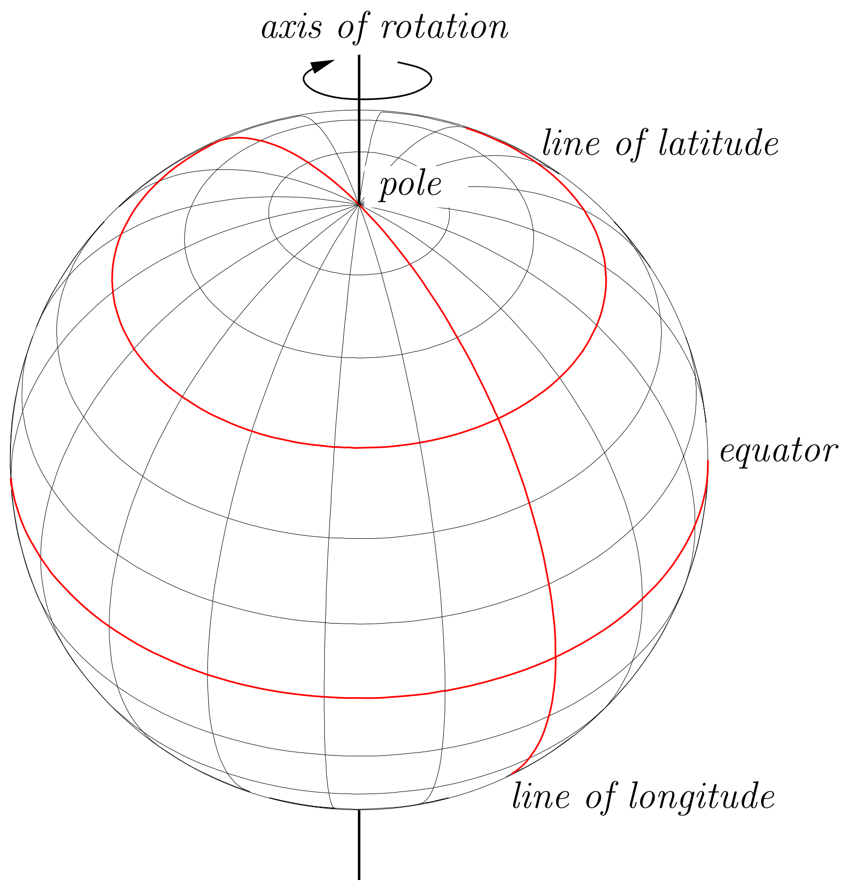


Figure 4.1: Poles, equator, lines of latitude and longitude of a sphere with a single axis of rotation

axis is aligned with the axis of rotation and the sphere is centered at the origin; thus the equatorial plane, given by $z = 0$, contains the x and y axes.

Any curve given by the intersection of a plane passing through the centre of the sphere with the sphere's surface is referred to as a *great circle*; the equator and all lines of longitude are great circles. Great circles are important in navigation as the shortest path on the surface of the earth between two points follows the great circle between those two points given by the intersection of the plane determined by those two points and the earth's centre with the earth's surface.

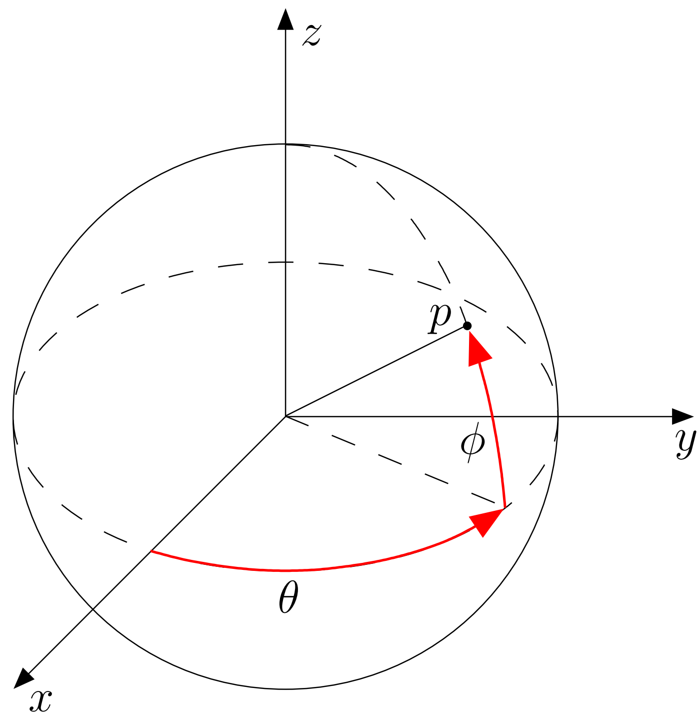


Figure 4.2: Radial spherical coordinates

4.1.1 Coordinate Systems

In three dimensional space the familiar Cartesian coordinate system provides three mutually orthogonal axes x, y and z, which locate any point by three coordinates giving the distance of the point from the origin along each axis. However for our purposes this is often more than we need; the surface of a unit sphere centred at the origin (with equation $x^2 + y^2 + z^2 = 1$) can be referenced in a simpler fashion using radial coordinates. Assuming a unit radius sphere, in radial coordinates a point p on the sphere can be located by two coordinates (θ, ϕ), where ϕ is the angle between the equatorial plane and the line between the origin and p, and θ is the angle between the positive x-axis and the line from the origin to p projected onto the equatorial plane (see Figure 4.2).

The spherical coordinates can be obtained from Cartesian coordinates, in the case of a unit sphere centered at the origin, by the transformation

$$\phi \;=\; \arcsin(z) \quad \text{and} \quad \theta \;=\; \arctan(\frac{y}{x}).$$

Using these spherical coordinates, any projection can be defined as a function

$$F : (\phi, \theta) \;\rightarrow\; (x, y)$$

that takes a point p on the 2-sphere onto a point p' on the Cartesian plane. Note that F may not be a bijection; in the case of the equirectangular cylindrical projection it maps the single point $(0, \theta)$ for any θ to the line $y = c$ for some constant c (chosen depending on the scale).

In n dimensions, the advantages of the spherical coordinate system is still apparent: a point is specified by a radial coordinate r and $n - 1$ angular coordinates $\phi_1, \phi_2, \ldots, \phi_{n-1}$; the unit sphere is specified by the equation $r = 1$.

4.1.2 Projections

Geometers refer to a function that maps the surface of the sphere (the domain), and by extension a diagram or figure on it, onto the Cartesian plane (the range) as a *map projection*, or just *projection* when the two surfaces involved are assumed from context. Various projections have been studied for centuries, due to their obvious applications in map-making, and most good atlases will have an introduction to the different projections used; see [120], or [98] for a good browseable online reference.

The surface of a sphere cannot be laid out on the plane without distortions, and so different projections have been developed that preserve different aspects of the spherical figure with various tradeoffs: projections exist that preserve area (Gall orthographic, sinusoidal, Albers conic), shape (conformal), direction to a fixed location (retroazimuthal), bearing or angle (Mercator, stereographic), distance from some standard point or line (equirectangular, Werner cordiform), and various "compromise" projections which seek to "make things look right" (Robinson, Fuller's Dymaxion, van der Grinten). In this work we will use only cylindrical projections and stereographic projections.

4.1.3 Cylindrical Projections

Cylindrical projections are possibly the simplest ways of mapping the surface of the sphere to a rectangle in the plane. They are characterized by an axis of rotation through the sphere that corresponds to the axis of rotation of a cylinder with the same diameter as the sphere; since we are usually dealing with diagrams with uniaxial rotational isometries we will always use cylindrical projections by a cylinder sharing that same axis; exceptions will be clearly noted.

In cylindrical projections lines of longitude on the sphere map to parallel vertical lines in the rectangle, the equator maps to a horizontal line of length 2π, and each pole maps to the entire line on the top (for the uppermost pole) or bottom (for the lower pole) of the rectangle; see the illustration[1] in Figure 4.3. All lines of latitude map to horizontal lines of equal length in the rectangular domain.

See Figure 4.4 for an example[2]; note the extreme distortion in area that occurs with objects near the poles.

Different cylindrical projections (Mercator, Miller, plate carrée, equal-area) differ in how they treat vertical distortion. In the simplest cylindrical projection, called the equirectangular projection, points on a line of longitude separated by an arc of given angle are always separated by the same distance vertically.

Given the spherical coordinate system, the equirectangular projection is given by the transformations $x = \theta$ and $y = \phi$; *i.e* the x-coordinate is given by the longitude and the y-coordinate by the arc distance from the equatorial plane. The two polar

[1]Figure redrawn from [49].
[2]Figure modified from [98].

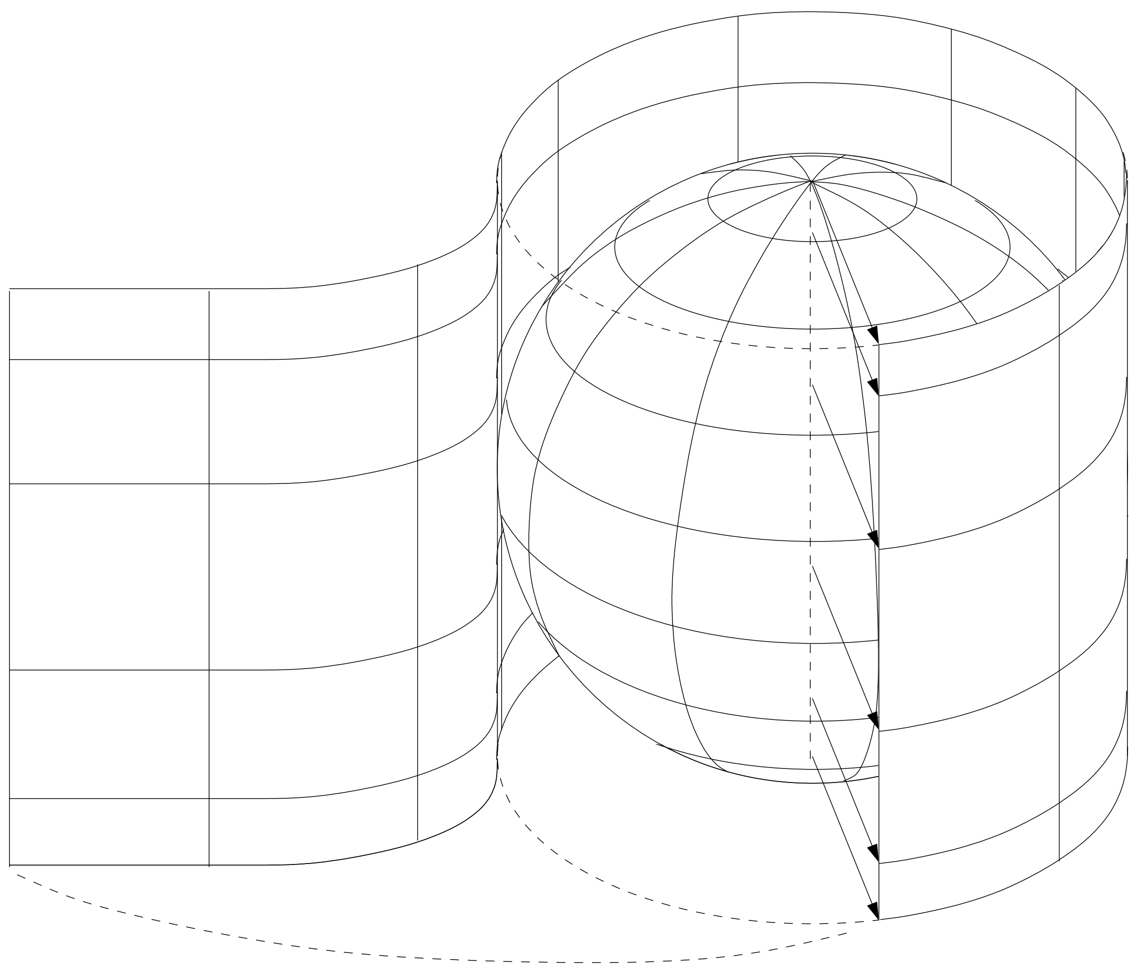


Figure 4.3: Cylindrical projection

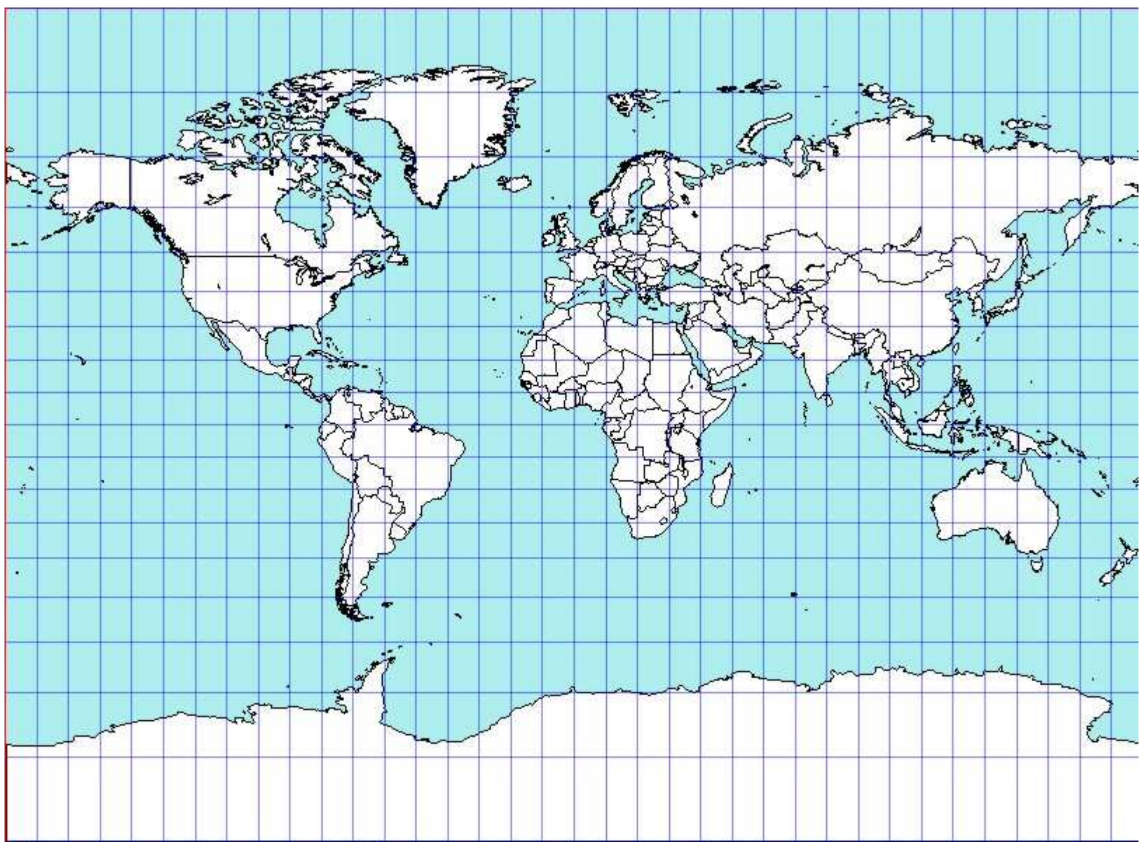


Figure 4.4: An example of a cylindrical projection, showing the sphere we are all most familiar with

points, at $\phi = \pm\pi/2$, map to the lines $y = \pm\pi/2$, which bound the rectangular domain on the top and bottom.

4.1.4 Stereographic Projection

For figures with multiple axes of symmetry or large, finite, symmetry groups, stereographic projections are simple and aesthetically pleasing representations, and are especially used in crystallography [74, Ch. 1]. An intuitive formulation is simple: given a diagram and a tangential plane, let the point common to the sphere and the plane be the *south pole* and the antipodal point be the *north pole*, which we shall refer to as the *projection point*, for reasons which will become clear. For any point p on the sphere, a line between the north pole (projection point) and p intersects the tangential plane at a point p', which is its image on the plane; the north pole maps to a point at infinity (in all directions) on the plane. Circles on the sphere map to circles on the plane, unless the circle passes through the north pole, in which case it maps to a line (which may be thought of as a circle through infinity). Also, stereographic projection preserves angles. It is useful to think of the projection as "unwrapping" the sphere from the south pole outwards onto the plane; see Figure 4.5.

See Figure 4.6 for an example[3], showing only the points north of the equator on a projection of the Earth to the plane tangent to the north pole, with the south pole as the point of projection.

Given our spherical coordinate system, the stereographic projection of a point (ϕ, θ) to the plane tangent to the south pole gives the point (x, y) with coordinates given by the formulae

$$x = 2\cos\theta/\tan(\pi/4 - \phi/2)$$

and

$$y = 2\sin\theta/\tan(\pi/4 - \phi/2).$$

[3]Figure modified from [98].

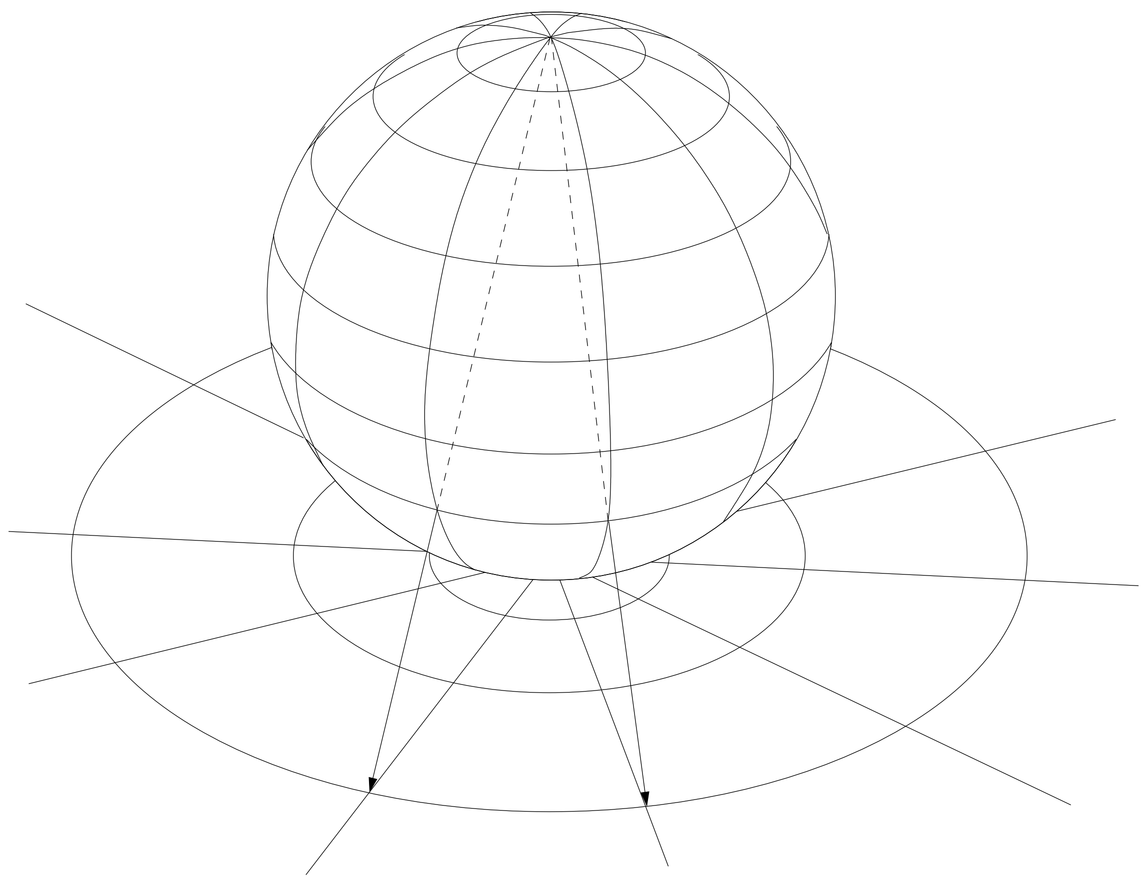


Figure 4.5: Stereographic projection of the sphere from the north pole onto the plane tangent to the south pole

It is often more convenient, however, to use polar coordinates on the plane, since then one coordinate (the angle θ) is preserved unchanged throughout the projection. Then the point (ϕ, θ) projects to the point (r, θ) on on the plane tangent to the south pole given by the (much simpler) formulae

$$r \;=\; 2\tan(\phi + \pi/4) \tag{4.1}$$

and

$$\theta \;=\; \theta.$$

The inverse projection (from the same plane back onto the sphere) is thus given by $\phi = \arctan(r/2) - \pi/4$, with θ preserved.

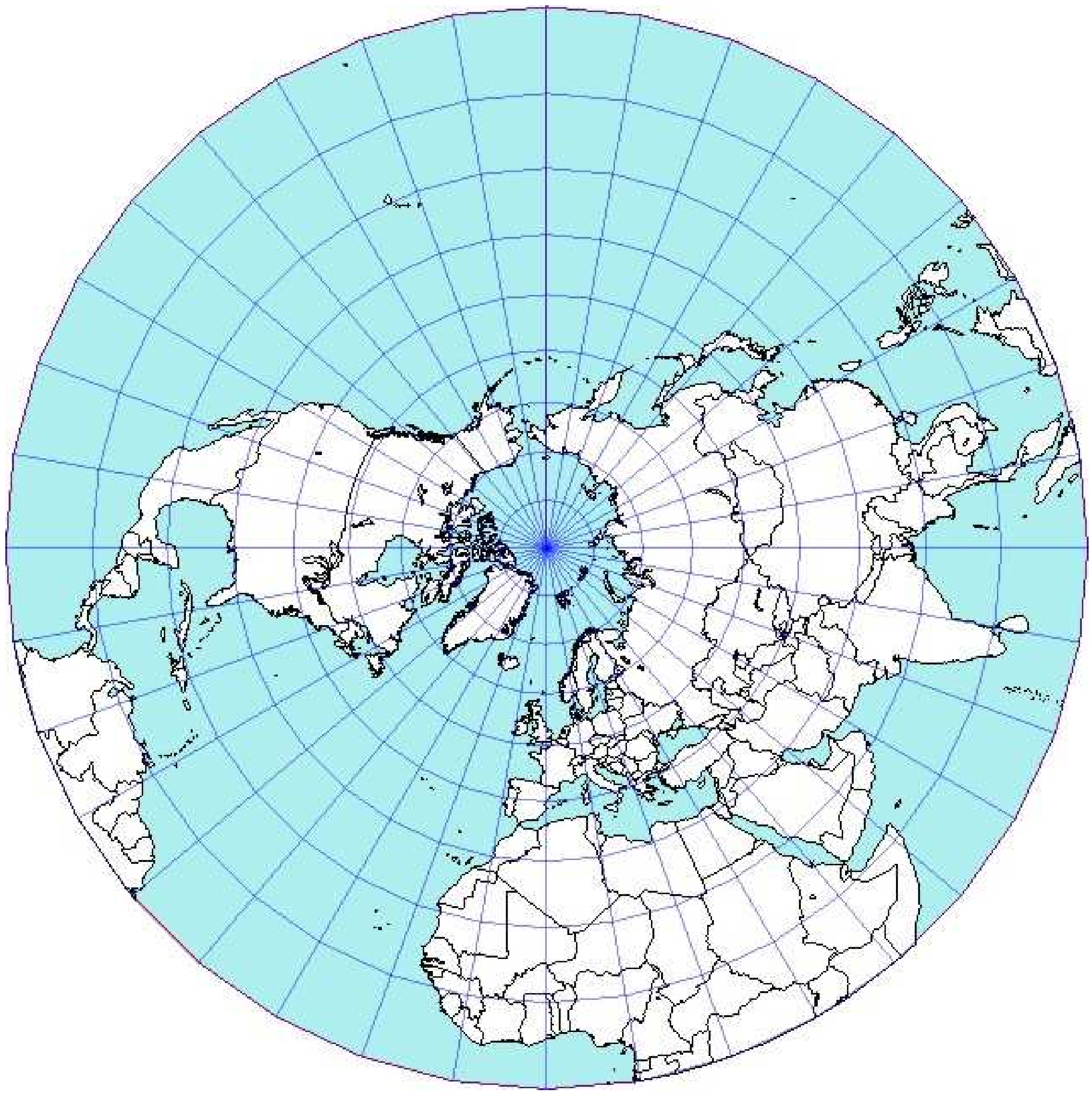


Figure 4.6: An example of stereographic projection, showing only the hemisphere nearest the tangent point, the northern hemisphere

When we are drawing stereographic projections of spherical diagrams, it is usual to choose the projection point such that it lies within a face of the spherical diagram: this face then becomes the (unbounded) outer face of the projected diagram, since it contains the image of the projection point at infinity.

4.2 Spherical Symmetries

In this section we will formalize some of our earlier, more intuitive, notions of symmetry with which we explored some of the previous research. The immediate consequences of these formalizations will lead to some properties and lemmas regarding diagrams on the sphere.

We have already seen the definition of isometries on the plane, and this definition naturally extends to other spaces. We have already discussed isometries of the sphere in the context of discussing the finite symmetry groups on the sphere.

Similar to our definitions of plane isometries in Section 2.1.3, we define the isometries geometrically. We assume the sphere in question is a unit sphere centered at the origin of 3-space. Since any isometry involving rotation has exactly one axis of rotation, and often a diagram has only one isometry involving rotation, we can assume unless otherwise stated that the axis of rotation coincides with the z axis (the vertical axis in Figure 4.2); thus the vector $\vec{v}$ defining this axis can be assumed to be $\vec{v} = (0, 0, 1)$.

This convention makes calculating the transformations simple and establishes a uniformity to our planar projections of spherical diagrams in the following way. For the cylindrical projections, the axis of rotation of any rotation-based isometry of a diagram on the sphere coincides with the axis of rotation of the cylinder used in the projection; the resulting planar figure will thus have a planar glide isometry under a horizontal shift, where points shifted horizontally past one edge of the diagram wrap to the other side of the diagram. For stereographic projections, the axis of rotation passes through the north and south poles of the diagram, and thus the resulting projection will exhibit a planar rotational isometry (as in Figure 2.6(b)) about the south pole.

The isometries of the sphere can be naturally grouped into three types:

1. rotation about an axis through the origin,

2. reflection across a plane through the origin, and

3. rotary reflection: a reflection across a plane through the origin followed by rotation about an axis orthogonal to the plane of reflection.

See Figure 4.7 for illustrations. These isometries are more formally defined as follows:

rotation: a rotation $R_{\vec{v},\psi}$ of a point p shifts p in the direction of ψ around the axis of rotation defined by $\vec{v}$, which is a vector giving the direction of the axis of rotation from the origin. Using our convention that the axis of rotation coincides with the vertical (z) axis, radial coordinates express the rotation of $p = (\phi, \theta)$ as $R_{\vec{v},\psi}(p) = (\phi,\ (\theta + \psi) \bmod 2\pi)$.

reflection: a reflection $F_{\vec{v}}$ of a point p reflects p across a plane through the origin orthogonal to the vector $\vec{v}$. In Cartesian coordinates, the plane orthogonal to $\vec{v} = (v_1, v_2, v_3)$ through the origin is all points $\vec{r} = (x, y, z)$ such that $\vec{v} \cdot \vec{r} = 0$, so the plane of reflection is given by $v_1x + v_2y + v_3z = 0$.

The full equation of the reflection of $p = (x_0, y_0, z_0)$ across the plane through the origin orthogonal to $\vec{v} = (v_0, v_1, v_2)$ is

$$F_{(v_0,v_1,v_2)}(p) \ = \ (x_0, y_0, z_0) - \frac{2(x_0v_0 + y_0v_1 + z_0v_2)}{v_0^2 + v_1^2 + v_2^2}(v_0, v_1, v_2).$$

Important special cases include reflection across the three coordinate planes, which has the effect of changing the sign of the coordinate not involved in that plane; for example, reflecting a point $p = (x, y, z)$ across the x, y plane (defined by $z = 0$) simply complements the z coordinate: the transformation is given by $F_{(0,0,1)}(p) = (x, y, -z)$.

rotary reflection: a rotary reflection $G_{\vec{v},\psi}$ is a combination of a reflection across the plane orthogonal to $\vec{v}$ through the origin and a rotation by ψ around the axis of rotation defined by $\vec{v}$. We use the notation $G_{\vec{v},\psi}$ so as not to confuse rotary reflections with rotations $R_{\vec{v},\psi}$, and because of the similarity of the operation to glide reflection in the plane; we will see later that a figure on the sphere with an rotary reflection isometry has a kind of glide reflection symmetry in its cylindrical projection onto the plane. Using our convention that the axis of rotation coincides with the vertical (z) axis, radial coordinates express the rotary reflection of $p = (\phi, \theta)$ as $G_{\vec{v},\psi}(p) = (-\phi,\ (\theta + \psi) \bmod 2\pi)$. Note that applying a rotary reflection twice gives a rotation, *i.e* $G_{\vec{v},\psi}(G_{\vec{v},\psi}) = R_{\vec{v},2\psi}$.

Some special cases are of interest:

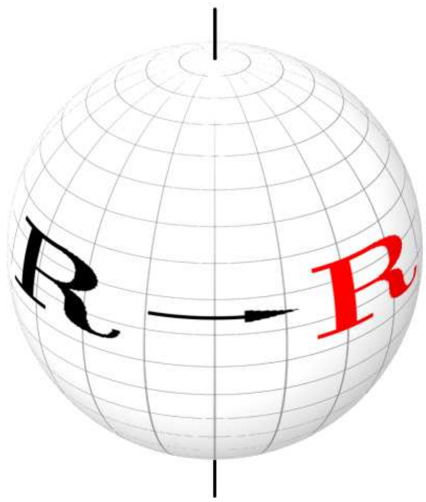
identity: *i.e* no transformation, which is equivalent to rotation of 2π,

inversion: equivalent to a rotary reflection $G_{\vec{v},\pi}$, where the rotation is of π (halfway around) followed by reflection, and

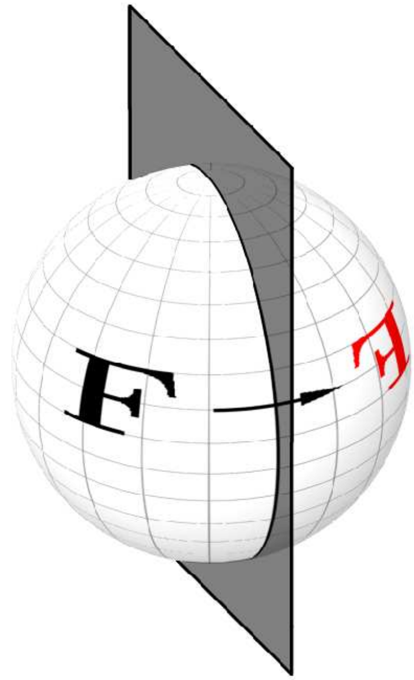
rotation by π: equivalent to two reflections across perpendicular planes.

Any of these operations transform any diagram embedded on the sphere to another diagram on the same surface. We modify our previous terminology as follows:

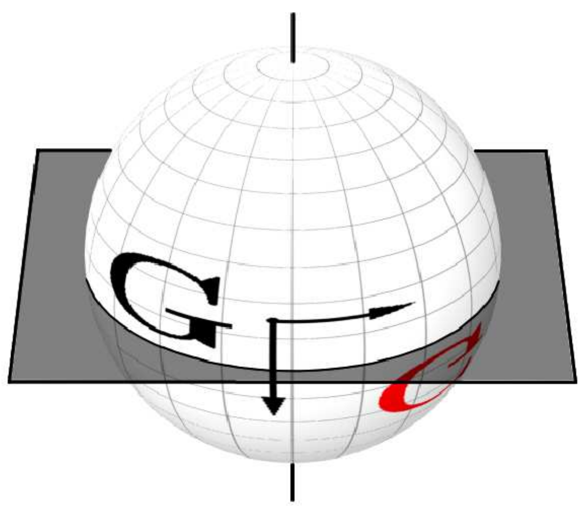
Definition. Given a diagram $D = (V, E)$ on a surface S, any transformation transforms D to a diagram $D' = (V', E')$; if $D = D'$ (modulo any labelling function), then the transformation is called a *symmetry* of D.



(a) Rotation about an axis through the origin



(b) Reflection across a plane through the origin



(c) Rotary reflection across a plane through the origin
followed by rotation around an orthogonal axis

Figure 4.7: Isometries of the sphere

Thus, for every point p on S,

- p maps onto a vertex $v' \in V'$ iff it is on a vertex $v \in V$ of D,
- p maps onto an edge $e' \in E'$ iff it is on an edge $e \in E$ of D, and
- p maps onto neither an edge nor vertex otherwise.

So a symmetry maps vertices onto vertices and edges onto edges.

A symmetry may preserve or change the labelling of D, according to the colour symmetry involved, which we discuss in the next section.

4.3 Colour Symmetry

We now consider some of the different types of symmetries possible for diagrams on the sphere and develop notation and initial properties for them. The differences between symmetries are possible because we are considering labelled diagrams: recall that we can interpret diagrams as edge-labelled graphs, where the label of an edge is the index of the curve it forms a part of. Since a symmetry of a diagram maps edges onto edges, the different types arise once we consider the effect the symmetry has on the edge-labelling.

The theory of colour symmetry is a relatively new branch in group theory, and has been developed, mainly by physicists, from the group theory of crystallography and diffraction and has been studied in the context of art (a good introduction is [103], see also [77, 87, 112]). As already noted, the symmetries of an object must form a group. Given an object realizing a given symmetry group, the theory starts from colouring the different fundamental regions of the object, and then examining the effects of applying various symmetries to the permutations of the colours generated.

Recall from Chapter 2 that the group of all isometries on the sphere is $O(3)$.

The first kind of symmetry is the most restrictive; it occurs when a symmetry only maps curves in the diagram back onto themselves; *i.e* edge labels are preserved.

Definition. A symmetry f of $D = \{C_1, C_2, \ldots, C_n\}$ is *total* if $f(C_i) = C_i$ for $1 \leq i \leq n$.

See Figure 4.8 for an example.

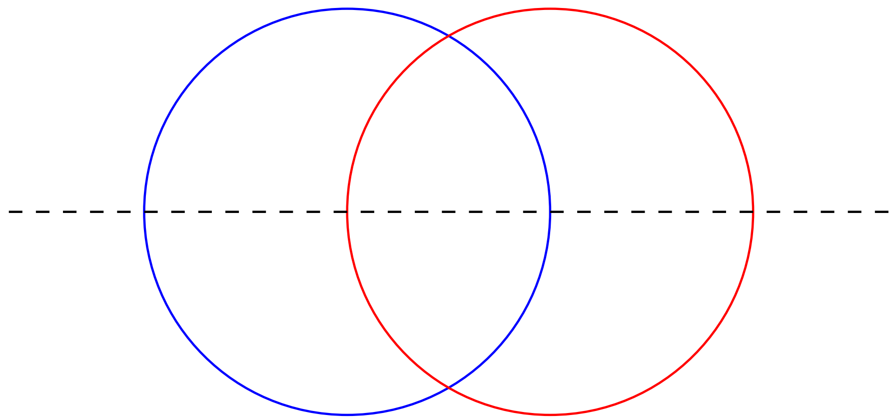


Figure 4.8: An example of total symmetry in a diagram; reflecting the 2-Venn diagram across the horizontal axis maps each curve onto itself

Given a diagram $D = \{C_1, C_2, \ldots, C_n\}$, let the set of symmetries (isometries and compositions of isometries) of D be $S(D)$. Then the group of total symmetries for D is denoted

$$S_T(D) = \quad \{\phi \in S(D) : \phi(C_i) = C_i,\ i = \{1, 2, \ldots, n\}\}.$$

A further generalization is when curves must map onto other curves in their entirety; thus, all edges of the same colour must all map onto edges of one other colour, producing an equivalent colouring. The diagram is thus invariant after applying the symmetry once the colours are relabelled appropriately.

Recall that the set of permutations of the n-set is written $\mathbb{P}_n$.

Definition. A symmetry f of $D = \{C_1, C_2, \ldots, C_n\}$ is *curve-preserving* if there exists a permutation $\pi \in \mathbb{P}_n$ such that $f(C_i) = C_{\pi(i)}$ for $1 \leq i \leq n$.

We have seen many examples already of diagrams with curve-preserving symmetries; see for example Figure 3.1. Curve-preserving symmetries are also referred to as *colour symmetries* in the literature [103], when examined in the context of general symmetry groups of coloured symmetric objects.

We denote the curve-preserving symmetry group of a diagram D as

$$S_C(D) = \{\phi \in S(D) : \exists \pi \in \mathbb{P}_n : \phi(C_i) = C_{\pi(i)},\ i = \{1, 2, \ldots, n\}\}.$$

In topology the concept of curve orientations is common, whereby a simple closed curve is given an orientation with respect to a region (on the plane, the exterior region). In later sections we will consider different variations of curve-preserving symmetries in which the orientations of curves are taken into consideration, but until then we assume that orientation of specific curves is irrelevant unless otherwise specified.

Finally, a symmetry can ignore colours (the labelling of the diagram) altogether.

Definition. A symmetry f of $D = \{C_1, C_2, \ldots, C_n\}$ is *monochrome* if $f(D) = D$; that is, it is not necessarily total or curve-preserving.

See Figure 4.9 for an example of a diagram with several reflective monochrome symmetries but no curve-preserving or total symmetries. This example, from [107], has a relaxed variant of the curve-preserving rotational symmetry on the plane discussed in Section 3.1. This particular variety of monochrome symmetry is referred to as *pseudo-symmetry*; see [107] for definitions and more examples, including several eight-curve pseudosymmetric Venn diagrams and a nine-curve diagram.

We denote the monochrome symmetry group of a diagram D as

$$S_M(D) = S(D),\ i.e \text{ the symmetry group of } D \text{ if all symmetries are allowed.}$$

4.3.1 Properties of Colour Symmetry Groups

We now consider some basic properties of the groups just defined.

Lemma 7. *Given a diagram D embedded on a surface,*

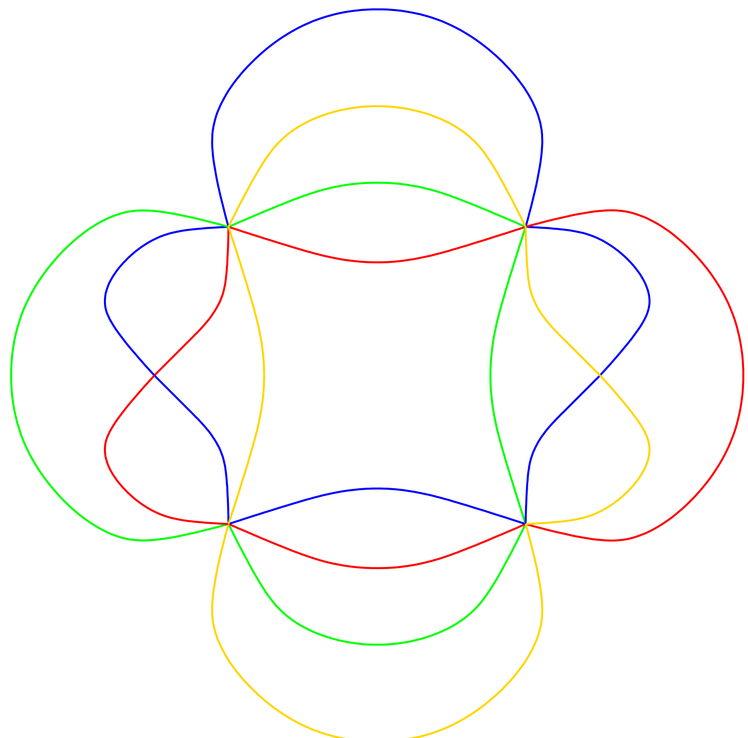


Figure 4.9: An example of monochrome symmetry in a 4-Venn diagram; this diagram has reflective symmetry across the horizontal and vertical axes, but none of them are curve-preserving

$$S_T(D) \ \leq \ S_C(D) \ \leq \ S_M(D) \ .$$

Proof. Any symmetry operation that results in a total symmetry is also a curve-preserving symmetry in which the colours are permuted by the identity permutation; thus, this set of symmetries forms a subgroup of all curve-preserving symmetries. Similarly, any symmetry operation that results in a curve-preserving symmetry also results in a monochrome symmetry. □

The following lemma provides a rather obvious way to test for the existence of curve-preserving symmetry operations that are not total symmetries.

Lemma 8. *Let $D = \{C_1, C_2, \ldots, C_n\}$ be a diagram. If there do not exist $C_i, C_j \in D$ with $i \neq j$ and an isometry f such that $f(C_i) = C_j$, then*

$$S_T(D) \ = \ S_C(D) \ .$$

Proof. Since no curve can map to any other curve but itself, even if curve-preserving symmetries are allowed, then the only symmetries possible can only map each curve to itself, which gives exactly the symmetry group under total symmetry. □

Theorem 4. *The (unique) two-curve Venn diagram is the only Venn diagram of at least two curves with a non-trivial (non-identity) total symmetry on the plane.*

Proof. Let D be a plane Venn diagram with a non-trivial total symmetry. Recall that a Venn diagram of n curves has 2^n regions. Let $f \in S_T(D)$ be a total symmetry; since f preserves curves and thus regions, the dual graph D' of D is preserved under f also, assuming that the vertex in D' corresponding to the external face in D is located at infinity and edges are drawn appropriately. The labels of vertices in D' are also preserved under f, and the vertex labels are unique in D' since D is a Venn diagram. Thus, every vertex of D' maps onto itself, and f is an automorphism of D'. The only non-identity isometry preserving every vertex of a graph on the plane with at least two vertices is a reflection applied to a graph with all vertices arranged on a line coincident with the line of reflection. The only Venn diagrams with their dual satisfying this are the (unique) one-curve diagram and (unique) two-curve diagram in Figure 4.8, since any diagram with $n > 2$ curves has at least $2^3 = 8$ vertices in the dual, which must be two-connected, and thus cannot be arranged on a line. □

Lemma 9. *Let D be a diagram. If D is simple and all of its curves intersect transversally, then*

$$S_C(D) \ = \ S_M(D) \ .$$

Proof. Under the operation of any monochrome symmetry $f \in S_M(D)$ on a diagram D, edges map to edges. If the diagram is simple, all vertices have degree four.

It is known that colouring algorithms for 4-regular graphs that are embeddings of diagrams produce a unique colouring once a single colour of an edge on each curve is specified. See, for example, [84, Proposition 2.2], which colours curves according to the 4-regular graph colouring algorithm of Iwamoto and Toussaint [75]. The proof follows this idea.

Since f is an isometry, it is easy to see that applying f to a diagram D will retain the property that all curves in D intersect transversally, since the order of colours encountered in a clockwise traversal around a given vertex will not change under the action of f.

Given f, edge $e \in C_i$ maps to edge $f(e)$ under f, and thus $label(e)$ is mapped to $label(f(e))$. The neighbouring edges to $e \in C_i$, *i.e* the edges in C_i that are adjacent to the endpoints of e, are the "opposite" edges to e across the neighbouring vertices: they are encountered third in a clockwise walk around either endpoint of e, starting at $e \in C_i$. These neighbouring edges map to neighbouring edges of $f(e)$ under f, and since $f(D)$ retains the property that all curves intersect transversally, then the neighbouring edges to $f(e)$ have the label $label(f(e))$ as well. See Figure 4.10 for an illustration.

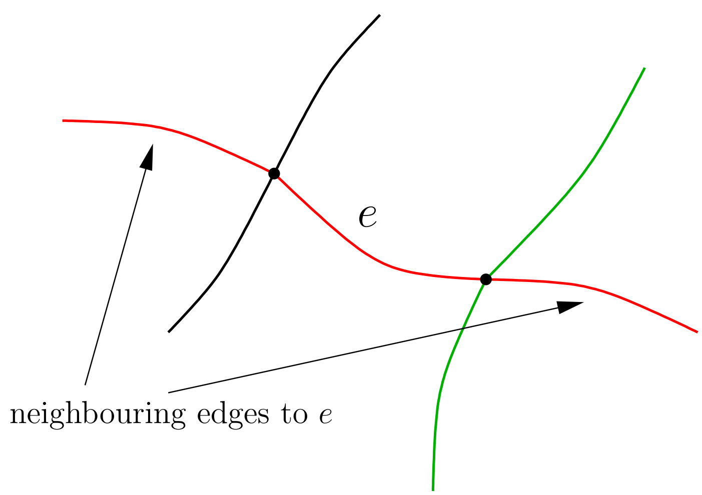


Figure 4.10: "Opposite" neighbouring edges to a given edge in a simple diagram must have the same colour

Continuing to follow the curve that $f(e)$ belongs to, f must map the entire curve of a given colour $label(e)$ to another entire curve of colour $label(f(e))$, and so the symmetry is also curve-preserving. Essentially, once the colour of a given edge is fixed the entire curve of that colour is fixed and can be recovered by simply following the edges across vertices, starting from the initial coloured edge. $\square$

We note that it appears that this lemma can be strengthened by relaxing the requirement that D be simple; it appears to be sufficient that all curves intersect pairwise transversally (*i.e* at every vertex, each pair of curves intersect transversally).

Finally, if a diagram D is a simple Venn diagram, its curves must intersect transversally, so we have the following Corollary:

Corollary 1. *Let D be a Venn diagram. If D is simple, then*

$$S_C(D) \ = \ S_M(D) \ .$$

4.4 Diagrams on the Plane

Any finite diagram on the plane must realize a finite symmetry group [10], and thus its symmetries cannot include translations, as these all give rise to infinite symmetry groups (the frieze groups or wallpaper groups, mentioned in Section 2.5).

Recall from Section 2.5 that the only finite plane symmetry groups are isomorphic to the cyclic groups and the dihedral groups. We have discussed symmetries of diagrams on the plane in the context of cyclic rotations, and in fact most of what is known about plane symmetry is confined to this type of symmetry. Table 4.1 displays Venn diagrams on the plane which realize the different symmetry groups listed[4]. Recall the isometries R_θ and F, where R_θ is a rotation by θ about a centre point and F is a reflection.

[4]Figures are reproduced from [106].

Table 4.1: Planar symmetry groups realizable by Venn diagrams

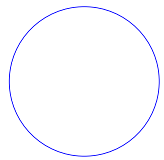
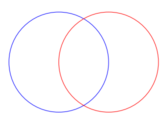
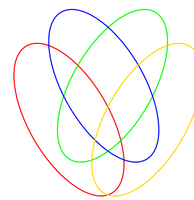
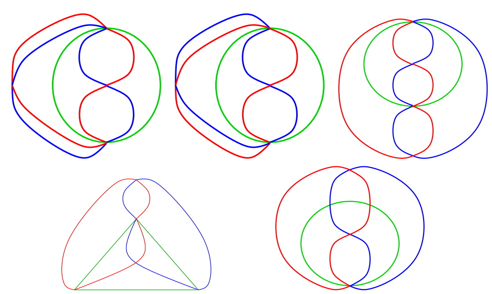
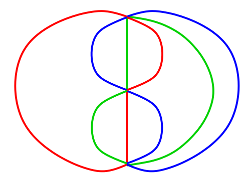
Isometry	**Group**	**Total Symmetry**	**Curve-preserving Symmetry**	**Monochrome Symmetry**
any R, F	C_m, D_m (any point group)	$n=1$: 		
F	C_2	$n=2$: 	$n = 4$:  $n = 3$: Fig. 3.3(b), 	$n=3$: 
R_π	C_2			$n = 8$: pseudosymmetric diagrams [106, p. "Variants on Symmetry"]

Table 4.1: (continued)

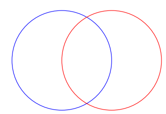
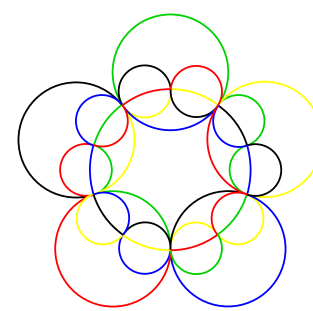
Isometry	Group	Total Symmetry	Curve-preserving Symmetry	Monochrome Symmetry
$R_{2\pi/3}$	C_3			$n = 9$: pseudosymmetric diagrams [ibid.]
$R_{2\pi/n}$	C_n		$n \geq 2$, prime: GKS construction (Section 3.3.2), several other 5-, 7-, and 11-Venn diagrams [106]	
F, orthogonal F	$\mathsf{D}_2 \cong \mathsf{C}_2 \times \mathsf{C}_2$		$n = 2$: 	$n = 4$: Fig. 4.9
F, $R_{2\pi/3}$	D_3		$n = 3$: Fig. 3.6(a)	$n = 3$: Fig. 3.6(b)
F, $R_{2\pi/5}$	D_5			$n = 5$: 

Table 4.1: (continued)

Isometry	**Group**	**Total Symmetry**	**Curve-preserving Symmetry**	**Monochrome Symmetry**
F, $R_{2\pi/7}$	D_7			$n = 7$: GKS construction [106, p. "Symmetric Diagrams, Necklaces, and Chains"]

4.5 Diagrams on the Sphere

As we noted in Section 2.1, any (finite) diagram on the plane can be embedded on the sphere. The definitions from Section 2.1 naturally generalize to the spherical case. The only interesting difference is that, on the plane, a Jordan curve divides the plane in to a (bounded) interior and (unbounded) exterior, and any point contained in one of these regions will remain so under continuous transformations of the curve. Given a Jordan curve C on the sphere, however, we can no longer distinguish the exterior region of C from the interior, as both are bounded: this allows the freedom to choose the exterior or interior however is convenient. The notions of regions, containment, nearness, transformations, and isometries all generalize naturally to the surface of the sphere; in Section 4.2 we saw how the planar isometries translate to the spherical case.

4.5.1 History

Some previous work has been done investigating diagrams, Venn and otherwise, embedded on the sphere. Grünbaum [58] introduced the notion of polar symmetry discussed in Section 3.2.2. He also introduced the notion that two diagrams are in the same *class* if one is isomorphic to the other on the sphere, and pointed out that the two unique simple 4-Venn diagrams on the plane are in the same class. The authors in [17] call a diagram on the sphere *spherical* (not to be confused with the use of the same term in [59] to refer to polar symmetric diagrams), and enumerate all simple 5-Venn diagrams on the sphere.

4.5.2 Oriented Symmetries on the Sphere

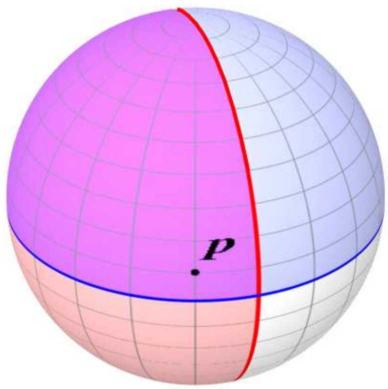
The flexibility to choose the interior side of each curve on the sphere discussed in the previous section applies to symmetries as well. On the plane, a curve-preserving symmetry must map the interior of each curve onto the interior, since a bounded region cannot be mapped to an unbounded region (or vice versa), however on the sphere no such restriction applies. Thus, a curve-preserving symmetry can map the interior of curve C onto the exterior of C (and vice versa). We can formally distinguish between two types of symmetry operation as follows: those that preserve "interiorness" (or exteriorness) and those that do not. We encapsulate this notion in the following definition. The idea is that a symmetry preserves "orientation" if, given an assignment of interiors and exteriors to all the curves, then the symmetry operation maps interiors onto interiors (and exteriors onto exteriors).

Definition. Let $D = \{C_1, C_2, \ldots, C_n\}$ be a diagram on the sphere, with curve-preserving symmetry group $S_C(D)$. For each C_i, specify an assignment of $\{$*interior*, *exterior*$\}$ to the partitions of the sphere by C_i. A symmetry operation $f \in S_C(D)$ is *oriented* if for $C_i \in D$, $f(\mathit{interior}(C_i)) = \mathit{interior}(f(C_i))$, (and thus $f(\mathit{exterior}(C_i) = \mathit{exterior}(f(C_i)))$, for $1 \leq i \leq n$. Any $f \in S_C(D)$ that is not oriented is *unoriented.*

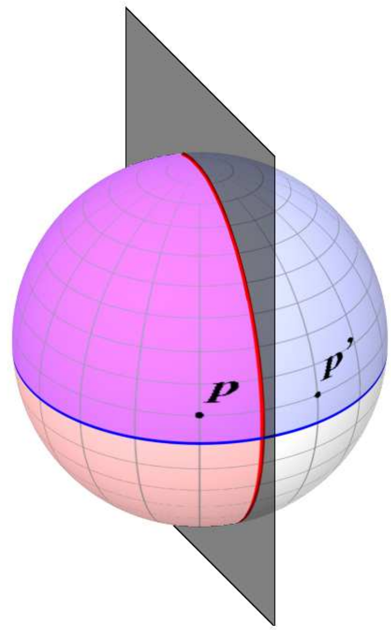
Since monochrome symmetry operations ignore curve identities, we restrict our definition above to only curve-preserving and total symmetries. For total symmetries, the above definition has the property that $f(C_i) = C_i$ for the curve mappings under f. The lemmas in Section 4.3.1 all refer to unoriented colour symmetries, since no orientation was specified for any curves on surfaces.

Example. Consider the diagram in Figure 4.11, consisting of two great circles, the blue equatorial curve and the red curve passing over the poles; let C_R be the red curve and C_B the blue curve. Furthermore, an assignment of interiors and exteriors is illustrated by the shading: $\mathit{interior}(C_R) \cap \mathit{exterior}(C_B)$ is shown in light red, $\mathit{interior}(C_B) \cap \mathit{exterior}(C_R)$ is light blue, and $\mathit{interior}(C_R) \cap \mathit{interior}(C_B)$, the intersection of the two curve interiors, is magenta. Also a point $p \in \mathit{interior}(C_R) \cap \mathit{interior}(C_B)$ is shown.

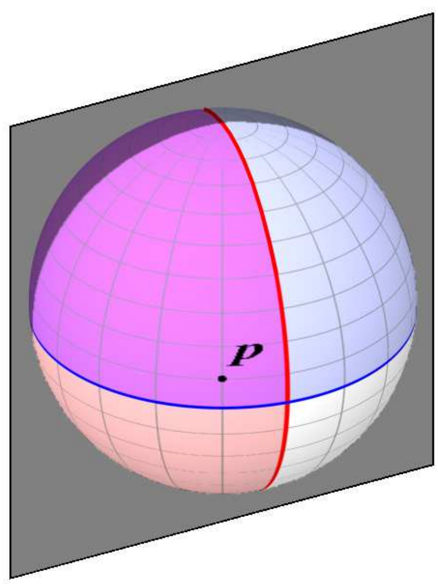
Consider the total reflective symmetry f, illustrated in Figure 4.11(b), which is a reflection across the plane containing C_R, and let $p' = f(p)$. Then f takes p



(a) Two-curve diagram with an assignment of interior/exterior to curves



(b) Vertical plane of reflection including the red curve



(c) Reflection across a vertical plane orthogonal to both curves

Figure 4.11: Examples of oriented versus unoriented symmetries

across C_R, and thus $p' \notin interior(C_R)$, or $p' \in exterior(C_R)$, and this is true for any point in $interior(C_R)$, since C_R is contained in the plane of reflection, and thus $f(interior(C_R)) = exterior(C_R)$, and thus f is unoriented, since it is clear that if the assignment of interior/exterior to the partitions of the sphere by C_R was reversed a similar situation would apply.

Now consider the total reflective symmetry f', illustrated in Figure 4.11(c), which is a reflection across a plane orthogonal to both C_R and C_B; using Cartesian coordinates, if C_B is embedded in the $x-y$ plane and C_R in the $y-z$ plane, then $f' = F_{\vec{c}}$ with $\vec{c}$ contained in the y axis, so f' is a reflection across the $x-z$ plane. Then f' reflects p to a point (not visible in the figure) $f(p) \in interior(C_R) \cap interior(C_B)$, *i.e* the point $f(p)$ remains in the same region. This applies for any point in this region and it is clear, by a similar argument, that all such interior/exterior regions are preserved (see the proof of Lemma 25, page 25, for a more formal version of a very similar argument). Thus the symmetry f' is oriented.

4.5.3 Oriented Total Symmetries

Oriented total symmetries are the most restrictive type of symmetries we consider. The following definition will be helpful in distinguishing when oriented total symmetries have non-trivial order.

Definition. Given a diagram D with symmetry group $S(D)$, an isometry $f \in S(D)$ is *region-preserving* if, for all regions R, any point $x \in R$ has the property that $f(x) \in R$; *i.e* under the isometry f, all points in R stay within that region.

Region-preserving symmetries are quite rare; as an example, consider the one-curve diagram in Figure 4.12 with a single equatorial curve: any uniaxial symmetry with a polar axis of rotation is region-preserving, including the rotation $R_{c,2\pi/n}$ giving a cyclic group of order n, for any n. This example when stereographically projected onto the plane tangent to a pole also has region-preserving symmetries, as the curve projects to a circle and the axis of rotation is the centre of the circle.

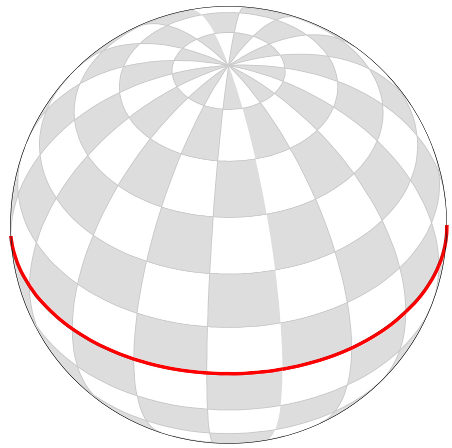


Figure 4.12: One-curve diagram with region-preserving symmetries

The following lemma provides a perspective on total symmetries on the sphere; by considering curve orientations we obtain the following lemma.

Lemma 10. *Unoriented total symmetries that are not region-preserving on the sphere are generators with orbits of order two.*

Proof. Given a diagram D on the sphere, the only symmetries possible are reflections, rotations, and compositions of the two. Reflections trivially have order two, so only rotations need to be considered. We consider two cases for the location of the axis of rotation.

Case 1: the axis of rotation passes through a region. Then after applying the symmetry, a point arbitrarily close to the axis of rotation remains in the same region, contradicting that the diagram has no region-preserving symmetries.

Case 2: the axis of rotation passes through at least one curve C. Let p be the vertex on C where the axis of symmetry intersects C, and think of p as a vertex. Since C is simple there are exactly two edges in C incident to p. The only configuration of these two edges that allows a total symmetry is for them to be embedded 180° apart, and the only rotation symmetry possible preserving these two edges is a 180° rotation, which has order two.

Finally, the composition of any reflections and rotations of order two will also have order two, completing the proof. □

Since all symmetries are order two, we know that the total group must be isomorphic to C_2^k, the composition of k cyclic groups of order two, for some $k > 0$. Several of the groups listed in Table 2.2, on page 43, are isomorphic to C_2^k, so we know which class the resulting group must be in. Thus we have the following corollary:

Corollary 2. *Let D be an n-curve diagram on the 2-sphere with at least one non-trivial unoriented total symmetry and no region-preserving unoriented total symmetries. Then* $S_T(D) \in \{\mathsf{C}_2, \mathsf{C}_{1h}, \mathsf{C}_{2h}, \mathsf{C}_{2v}, \mathsf{D}_2, \mathsf{D}_{1h}, \mathsf{D}_{2h}, \mathsf{D}_{1d}, \mathsf{S}_2\}$.

Note that some of the classes in Table 2.2 overlap when the parameter n is small, as it is in these examples, and some of the groups listed are different notations for the same group (with the same isometries); we list them all for completeness.

Also recall from Section 2.4.4 that the integers $\mathbb{Z}_n = \{0, \ldots, n-1\}$ under addition modulo n form a cyclic group of order n. Then the total unoriented symmetry group of a diagram D is isomorphic to $\mathbb{Z}_2^k$ for some $k > 0$.

Since oriented total symmetries are not as interesting, in this work we will always consider total symmetries to be unoriented, unless indicated otherwise.

4.5.4 Oriented Curve-Preserving Symmetries

Depending on if we are using oriented or unoriented symmetry operations, we can extend our original notation as follows. Given a diagram D, we refer to

$\mathbf{S_{OC}(D)}$ $= \{f : f \in S(D)$ and f is oriented$\}$, is the set of *oriented* curve-preserving symmetries of D.

Lemma 11. *$S_{OC}(D)$ is a group, under composition of isometries.*

Proof. Given a diagram D with unoriented curve-preserving symmetry group $S_C(D)$ with some assignment of interiors and exteriors to the curves, we modify D to give a new diagram D' such that $S_C(D') = S_{OC}(D)$. Since $S_C(D')$ must form a group, then $S_{OC}(D)$ forms a group.

Given $D = \{C_1, C_2, \ldots, C_n\}$, create an assignment of $\{interior, exterior\}$ to the partitions of the surface of the sphere for each C_i. Let D' be the same as D, with the following additions. Choose a point p on one curve C_i (and no others, so p is not a vertex of D). At p add an arbitrarily-small widget shaped like a carat (“$\wedge$”), with the point of the carat directed towards the interior of C_i. Then, for all $f \in S_C(D)$, add a widget at the location of $f(p)$, pointing towards the interior of $f(C_i)$, unless there is already a widget at $f(p)$ pointing in the opposite direction, or a widget already on the curve $f(C_i)$. If any curve does not have a widget added by this process, continue the process over all curves C_i, so that all curves have their orientation “specified” in the curve structure themselves.

Now consider the group $S_C(D')$. Any $f \in S_C(D')$ must preserve the widgets on the curves, and so f preserves the orientation encoded with the widgets, and so $S_C(D') = S_{OC}(D)$, and $S_{OC}(D)$ is a group. $\square$

The following corollary follows from the proof of the above lemma.

Corollary 3. *Given a diagram D, $S_C(D) \geq S_{OC}(D)$.*

Since we are assuming that total symmetries are always unoriented unless otherwise specified, for consistency, we always assume that curve-preserving symmetries are unoriented, unless otherwise specified.

Observation 1. *Let D be a diagram with $S_C(D)$ its group of unoriented curve-preserving symmetry operations on D, with an assignment of interior/exterior to the* *partitions of the sphere by each curve. If, for all* $f \in S_C(D)$*,* f *is oriented, then* $S_{OC}(D) = S_C(D)$*.*

Observation 1 clearly applies for all diagrams in the plane. On the sphere, if a symmetry does not change orientation of any curve mapping to any other curve (that is, if $x \in \mathit{interior}(C_i)$ implies $f(x) \in \mathit{interior}(f(C_i))$ for all x and C_i), then Observation 1 also applies.

4.5.5 Planar Symmetric Diagrams Embedded on the Sphere

Chapter 3 discussed planar symmetric Venn diagrams, including the necessary condition that n must be prime for a planar symmetric n-Venn diagram to exist, and the GKS construction (from [54]) that showed that this condition was sufficient as well. In this section we discuss how this planar symmetry translates to spherical symmetry.

Consider a planar symmetric diagram D, and embed it on the sphere to obtain the diagram D' via a stereographic projection, identifying the point of rotation on the plane with a pole on the sphere. The axis of rotation passing through the two poles on the sphere thus is an axis of symmetry for D'. This symmetry is curve-preserving, by the definition of planar symmetry, and since on the plane the interior and exterior of each curve is preserved under the symmetry, the symmetry is oriented. Thus we have the following.

Lemma 12. *Given a planar symmetric diagram* D *of* n *curves, a diagram* D' *of* n *curves on the sphere can be constructed so that* $S_{OC}(D') = \mathsf{C}_n$*.*

Proof. By the previous discussion, and Observation 1 and discussion thereof in the preceding section. □

Nearly all of the previous work in symmetry in diagrams has been of this type, and so we already know of many n-Venn diagrams with oriented curve-preserving symmetry group C_n on the sphere.

4.5.6 Polar Symmetry on the Sphere

In Section 3.2.2 *polar symmetry* of planar diagrams was introduced. In this section we discuss an application of polar symmetry of diagrams on the sphere by considering the spherical symmetry group of a given planar symmetric diagram with polar symmetry. As in Section 4.5.5, given a planar diagram D with polar symmetry, consider the stereographic projection D' of D onto the sphere about the point of rotation. By the definition of polar symmetry, the stereographic projection of D' back to the plane about the opposite point on the sphere gives a diagram D_P isomorphic to D. This isomorphism allows us to prove the following lemma.

Lemma 13. *Let D be a polar symmetric n-curve diagram on the plane, and let D' be the stereographic projection of D onto the sphere about the point of rotation. Then D' can be drawn so that $S_C(D') \geq \mathsf{D}_n$,* i.e *$D'$ has curve-preserving symmetry group equal to the polyditropic group, or includes it as a subgroup.*

Proof. Given a rotationally symmetric diagram D in the plane, let D' be its stereographic projection onto the unit sphere with the south pole tangential to the point of symmetry on the plane. Let D_P then be diagram formed by taking the inverse projection back onto the plane tangent to the north pole, and then taking mirror reflection across a line through the point tangent to the north pole. By definition of polar symmetry D_P is isomorphic to D, and observe that all curves in D are congruent to each other, by definition of rotational symmetry, and the curves in D_P are isomorphic to those in D, by definition of polar symmetry, so all curves in D_P are isomorphic to all curves in D.

We use polar coordinates on the plane to simplify the transformations involved. By the equations for stereographic projection and inverse stereographic projection in Section 4.1.4, we can compute the transformation of a point (r, θ) on the plane onto the unit sphere with its south pole tangent to the plane and then onto the plane tangent to the north pole to give the point (r_P, θ_P). By simply inverting the sign of the angle ϕ on the sphere we have the stereographic projection from the point (ϕ, θ') on the sphere onto the plane tangent to the north pole. This, combined with the projection onto the sphere of (r, θ), gives the point

$$\begin{aligned}
(r_P, \theta_P) &= (2\tan(-\phi + \pi/4), \theta) \qquad \text{by Equation 4.1} \\
&= (2\tan(-[\arctan(r/2) - \pi/4] + \pi/4), \theta) \qquad \text{by Equation 4.1} \\
&= (2\tan(\pi/2 - \arctan(r/2)), \theta) \\
&= (2\cot(\arctan(r/2)), \theta) \qquad \text{by the identity } \tan(\pi/2 - x) = \cot x \\
&= (4/r, \theta)\ .
\end{aligned}$$

This operation in the plane is sometimes referred to as *inversion* across a circle, in this case the circle $r = 4$. Reflecting (r_P, θ_P) across an arbitrary line, we can choose a line that is convenient, such as the line $\theta = 0, \pi$, to give the point $(r_P, \theta_P) = (4/r, 2\pi - \theta)$. The inversion followed by the reflection is referred to as the *polar symmetry operation*. The resulting diagram D_P will also have rotational symmetry and all its curves will be congruent.

We noted in Section 3.2.2 that the polar symmetry operation can map a curve back onto itself under the symmetry, but for this mapping to be preserved as an isometry it depends upon the scaling effect of the polar symmetry—for example, if $r > 4$ for all points in D, it is clearly not possible for a direct isometry to preserve curves from D to D_P. However, this can easily be remedied by either altering the radius of the sphere used as the intermediate step in the stereographic projection, or scaling diagram D. If each curve in D is drawn so that it is half inside and half outside the circle given by $r = 2$ (the circle which maps to the equator on the sphere), then this property will be preserved under the polar symmetry operation. More specifically, let the dual graph of D be constructed to be balanced about the circle $r = 2$ (with half of the nodes internal to each curve on either side of $r = 2$), and the diagram drawn so that an edge of the diagram inside of $r = 2$ will have the same shape and size of a corresponding edge outside of $r = 2$ once projected onto the sphere. In this way a scaling factor can be applied to preserve the size of all curves under the polar symmetry.

Example. For example, consider the 5-Venn diagram in Figure 3.1. This diagram is polar symmetric, though it may not be readily apparent. Figure 4.13 shows the cylindrical projection of the stereographic projection onto the sphere, with the equator shown as a dotted line. The equator thus is the projection of the circle $r = 2$ on the plane, and it is apparent that the structure of the diagram is drawn balanced across this line.

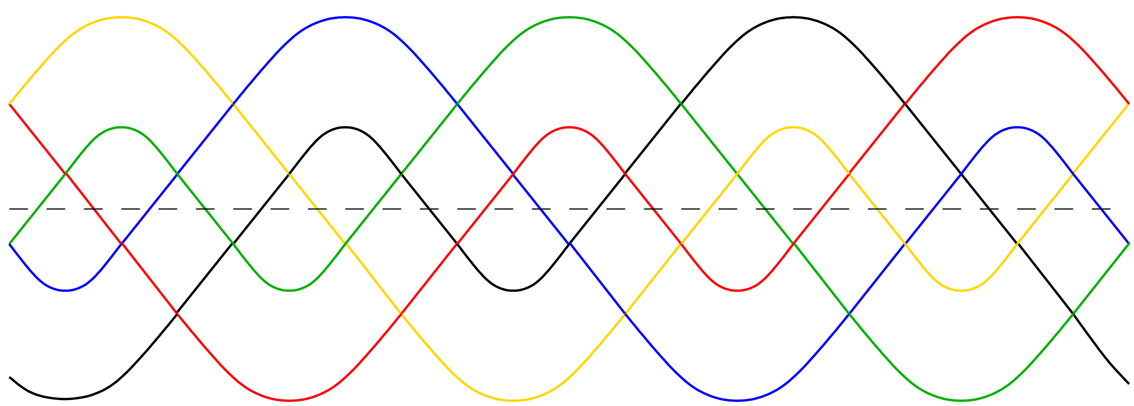


Figure 4.13: Cylindrical projection of Figure 3.1, with equatorial line shown dotted

Now consider the given operation of polar symmetry on the intermediate diagram D' on the sphere. As in Figure 4.13, the polar symmetry operation equates to a reflection across a horizontal line (such as the equator), given by the initial transformation $(r, \theta) \to (4/r, \theta)$, followed by a reflection across a vertical line, given by the subsequent transformation $\theta \to 2\pi - \theta$. The composition of these two operations is a rotation by π around some central point, written $R_{\vec{v},\pi}$, and the property of polar symmetry states that this operation results in an isomorphic diagram. We have argued that the structure of the diagram can be drawn such that edges and vertices are located equivalently with respect to the horizontal equator and thus this operation can result in an isometry, given appropriate choice of the axis of rotation. Thus $R_{\vec{v},\pi}$ is an isometry on D', given appropriate choice of $\vec{v}$.

These properties enable us to draw D so that the diagram D_P, ignoring colours, is isometric to D. Since both diagrams project to D' on the sphere, we have that if an isometry is performed to identify $C \in D_P$ with its image $C_P \in D_P$ then this isometry identifies $C'_i \in D'$ with some curve $C'_j \in D'$. Thus on D' this isometry induces some permutation of the curve colours on the remaining curves. Since all curves are congruent by the rotational symmetry, and the structure of the diagram is congruent under the polar symmetry, each curve in D' can be drawn so that its structure is congruent to all curves under the polar symmetry operation, and thus to itself under some axis of rotation.

Example. Continuing our example, Figure 4.14 shows the cylindrical projection of the polar symmetric 5-Venn diagram from Figure 4.13 with the axis of rotation $\vec{v}$ for the black curve passing through the points v and v'; this axis passes through the centre of the sphere and intersects the curve at two antipodal points on the equator. The rotation given by $R_{\vec{v},\pi}$(black) is thus an curve-preserving isometry and induces the colour permutation (black)(red green)(blue yellow) on the curves, and there are four other similar isometries, one for each curve.

Figure 4.15 shows the diagram from Figure 3.1 embedded on the sphere, with symmetry group D_5.

Thus each curve $C_i \in D'$ on the sphere has an associated isometry $R_{\vec{v},\pi}(i)$, with $1 \leq i \leq n$, with the axis of rotation $\vec{v}$ passing through the curve at two points where it crosses the equator, and the isometry induces a colour permutation on the curves, with the colour permutation fixing the colour of C_i since it maps to itself.

The diagram D' preserves the rotational symmetry $R_{\vec{v},2\pi/n}$ giving the rotation

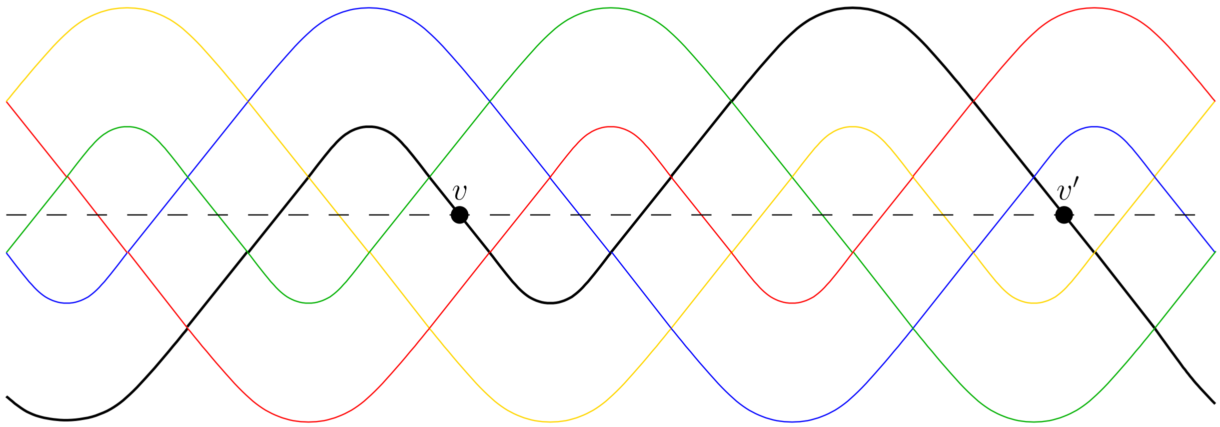


Figure 4.14: Figure 4.13, with the axis of rotation for the black curve passing through v and v'

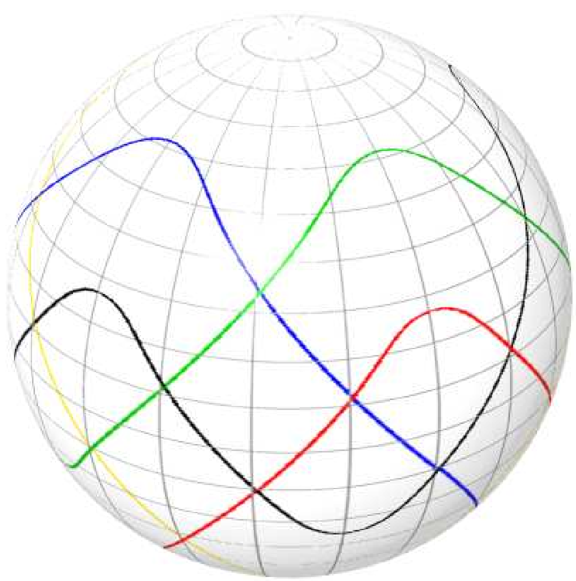


Figure 4.15: Stereographic projection of Figure 3.1 onto the sphere

group C_n about the polar axis of rotation; this is the rotation about the polar axis through the sphere that is present by definition of the rotational symmetry of D and is preserved through the stereographic projection to D'. This, combined with the n curve-preserving isometries $R_{\vec{v},2\pi/n}(i)$, $1 \leq i \leq n$, each of order two, gives the full dihedral rotation group. The diagram D' could have more symmetries, so we have that $S_C(D') \geq \mathsf{D}_n$, as desired. □

Lemma 13 immediately gives us a number of Venn diagrams on the sphere with rich symmetry groups, as there are many polar symmetric Venn diagrams already known in the plane, for prime $n \leq 7$; for several examples see the survey [106]. There are six simple monotone symmetric 7-Venn diagrams with polar symmetry, and thus these diagrams have curve-preserving symmetry group D_7, of order 14, when appropriately embedded on the sphere.

By reversing the construction in the proof of Lemma 13, we also have the following corollary:

Corollary 4. *Let D be an n-curve diagram on the sphere with curve-preserving symmetry group $S_T(D) = \mathsf{D}_n$, the polyditropic group, with the property that D has no curves passing through the axis of rotation at the poles. Then there is a stereographic projection D_P of D onto the plane tangent to sphere at a pole of rotation such that D_P is rotationally symmetric and has polar symmetry.*

Chapter 5

Symmetries in Chain Decompositions

In this chapter we explore the Greene Kleitman chain decomposition and investigate some different symmetries created by different variations of the decomposition and ways of embedding of the chains. We begin by providing a motivating example theorem, and then discuss some of the poset theory behind it.

Recall from Section 2.3.3 that the base case is the decomposition for $\mathcal{B}_1$, which is the single chain $\{0, 1\}$, and the recursive construction rule, in which every chain in a decomposition for $\mathcal{B}_n$, written $\sigma_k,\ \sigma_{k+1},\ \ldots,\ \sigma_{k+j}$, is replaced by the two chains

$$\sigma_{k+1}\,0,\ \ldots,\ \sigma_{k+j}\,0\ ,\quad \text{and}$$
$$\sigma_k\,0, \sigma_k\,1,\ \ldots,\ \sigma_{k+j-1}\,1,\ \sigma_{k+j}\,1\ ,$$

in the decomposition for the $\mathcal{B}_{n+1}$ lattice, and the first of these chains is omitted when $j = 0$.

The Greene Kleitman chain decomposition has several innate symmetries. The following theorem is implied by some other results in [83]; we provide a direct proof.

Theorem 5 ([83, ex. 76 p. 39])**.** *If* $\sigma_1, \sigma_2,\ \ldots\ , \sigma_k$ *is a chain in the Greene Kleitman chain decomposition, then so is*

$$\overline{\sigma_k}^R, \ldots, \overline{\sigma_2}^R, \overline{\sigma_1}^R\ ,$$

the reverse sequence of reverse complements.

Proof. Recall from Section 2.3.3 (Equation 2.1) that any binary string σ_i can be written in the form

$$\alpha_0\, 0\, \ldots\, \alpha_{p-1}\, 0\, \alpha_p\, 1\, \alpha_{p+1}\, \ldots\, 1\, \alpha_q$$

for some p and q with $0 \leq p \leq q$, where each substring α_i is a properly nested binary string, and possibly empty. In the Greene Kleitman pattern, σ_i is the $(q-p+1)$th element of the chain of length $q+1=k$,

$$\alpha_0\, 0\, \ldots\, \alpha_{q-1}\, 0\, \alpha_q,\ \alpha_0\, 0\, \ldots\, \alpha_{q-2}\, 0\, \alpha_{q-1}\, 1\, \alpha_q,\ \ \cdots\ \ ,\ \alpha_0\, 1\, \alpha_1\, \ldots\, 1\, \alpha_q\ .$$

Note that if α_j is a properly nested binary string, then $\overline{\alpha_j}^R$ is also.

The reverse complement string $\overline{\sigma_i}^R$ is

$$\overline{\alpha_q}^R\, 0\, \overline{\alpha_{q-1}}^R\, 0\, \ldots\, \overline{\alpha_{p+1}}^R\, 0\, \overline{\alpha_p}^R\, 1\, \overline{\alpha_{p-1}}^R\, \ldots\, 1\, \overline{\alpha_0}^R\ ,$$

and the reverse complement of the entire chain $\sigma_1, \sigma_2, \ldots, \sigma_i, \ldots, \sigma_k$ is then

$$\overline{\sigma_k}^R, \ldots, \overline{\sigma_i}^R, \ldots, \overline{\sigma_2}^R, \overline{\sigma_1}^R$$

which then expands to be

$$\overline{\alpha_q}^R\, 0\, \ldots\, \overline{\alpha_1}^R\, 0\, \overline{\alpha_0}^R,\ \ \cdots\ \ ,\ \overline{\alpha_q}^R\, 0\, \overline{\alpha_{q-1}}^R\, 1\, \overline{\alpha_{q-2}}^R\, 1\, \ldots\, 1\, \overline{\alpha_0}^R,\ \ \overline{\alpha_q}^R\, 1\, \overline{\alpha_{q-1}}^R\, 1\, \ldots\, 1\, \overline{\alpha_0}^R\ \ ,$$

with $\overline{\sigma_i}^R$ occurring at position $q+1-(q-p+1)=p$.

Since $\overline{\sigma_k}^R$ has no unmatched 1s it is a valid chain starter, and $\overline{\sigma_1}^R$ has no unmatched 0s and so is a valid terminator for that chain, and it is clear that applying the Greene Kleitman rule to this chain starter gives exactly the chain as shown, so the full chain will also occur in the decomposition. Note that the chains $\sigma_1, \sigma_2,\ \ldots\ , \sigma_k$

and $\overline{\sigma_k}^R, \ldots, \overline{\sigma_2}^R, \overline{\sigma_1}^R$ may in fact be the same chain, which will occur if $\overline{\sigma_k}^R = \sigma_1$. $\square$

This theorem indicates that diagrams created from the Greene Kleitman decomposition will have some symmetries already present, but these symmetries in fact are only manifested in the resulting diagrams as monochrome symmetries, which we show in Section 8.3. The following sections explore further symmetries that can be obtained by suitably modifying the Greene Kleitman decomposition.

This chapter is joint work with Frank Ruskey.

5.1 Symmetries in Posets

Recall our discussion of posets from Section 2.3.1; in this section we introduce some further terminology and concepts that we exploit later in the chapter for the purposes of building Venn diagrams. We continue to follow [123, 124] in our definitions.

Definition. When $P = (X, R)$ and $Q = (Y, S)$ are posets, a map $f : X \to Y$ is *order-preserving* (respectively, *order-reversing*) if $x_1 \prec x_2$ in P implies $f(x_1) \prec f(x_2)$ (respectively, $f(x_1) \succ f(x_2)$) in S for all $x_1, x_2 \in X$.

Definition. If two posets have an order-preserving (respectively, order-reversing) bijection between them, they are *isomorphic* (respectively, *dual-isomorphic*). An isomorphism from P to P is called an *automorphism*, and a dual-isomorphism from P to P is called a *dual-automorphism*. A poset with a dual-automorphism is termed *self-dual*.

Many of the classic posets studied in combinatorics are self-dual, including the boolean lattice.

The *dual* of a poset $P = (X, R)$ is the poset $P^* = (X, R^*)$ where $x \prec y$ in P if and only if $y \prec x$ in P^*; the Hasse diagram of P^* is the Hasse diagram of P drawn upside down.

We can extend these notions to chain decompositions, as some dual-automorphisms preserve the chains in a given chain decomposition.

Definition. Given two posets P and Q and chain decompositions C of P and E of Q, a *chain decomposition isomorphism* (respectively, *chain decomposition dual-isomorphism*) is an order-preserving (respectively, order-reversing) map $f : P \to Q$ such that $f(C_i) = E_j$ for some $C_i \in C$ and $E_j \in E$.

A chain decomposition isomorphism (respectively, dual-isomorphism) from a poset and chain decomposition P, C to P, C is a *chain decomposition automorphism* (respectively, *chain decomposition dual-automorphism*). Again, a chain decomposition that has a chain decomposition dual-automorphism is called *self-dual*.

Example. For example, in Figure 5.2 on page 113, let $f(x_1x_2x_3x_4) = \overline{x_1}\,\overline{x_2}\,\overline{x_3}\,\overline{x_4}$. Then f is a chain decomposition dual-automorphism on this chain decomposition of $\mathcal{B}_4$. Consider numbering the chains from top to bottom $1 \ldots 6$, then f also has the property that $f(C_i) = C_{7-i}$.

Theorem 5 in the previous section established the following lemma.

Lemma 14. *The Greene Kleitman chain decomposition of $\mathcal{B}_n$ is self-dual under the dual-automorphism $\phi : \phi(x) = \overline{x}^R$;*

5.2 Embeddings of Chain Decompositions

Some chain decompositions have the property that they form part of the dual graph of a Venn diagram on the plane; for this to be true there must be a planar embedding of the chains on the plane with appropriate additional edges so that they are linked into a connected graph. In Section 3.3.1 we discussed the construction of [54] in which chains in the symmetric chain decomposition of $\mathcal{B}_n$ were laid out on the plane in a specific order so that, once the chains were connected by chain cover edges and the dual graph taken, the dual formed a n-Venn diagram.

We begin by slightly modifying the appropriate definitions from Section 3.3.1. The concept of the chain cover property applying to a chain decomposition can be broadened to include non-symmetric chains, and allow chains to be covered by two different chains (one at each end).

Definition. A chain decomposition $\mathcal{C}$ has the *chain cover property* if whenever $C \in \mathcal{C}$, then there exists two chains $\pi(C), \pi'(C) \in \mathcal{C}$ such that

$starter(C)$ is covered by a starter element $\pi_s(C)$ of $\pi(C)$, and

$terminator(C)$ covers an element $\pi'_t(C)$ of $\pi'(C)$,

unless $starter(C)$ is of the smallest rank, in which case $\pi(C)$ does not exist, or $terminator(C)$ is of the largest rank, in which case $\pi'(C)$ does not exist. Again π is called the *chain cover mapping*.

For example, in Figure 5.2 the chain $\{0000, 0001, 0011, 0111\}$ has no chain covering its starter 0000, since 0000 is of minimal rank, but the terminator 0111 covers 1111, from the neighbouring chain. Figure 5.7 in Section 5.6 shows a different chain decomposition of $\mathcal{B}_4$ in which the chain $\{0110\}$ has two different chains covering its starter and terminator.

In [54], as outlined in Section 3.3, the concept of a chain decomposition is developed along with its embedding in the plane, via the concept of the chain cover graph. In this section we supersede this concept, since the use of the chain cover graph to embed the entire chain decomposition and all of its cover edges depend on the chain decomposition being symmetric, which is not the case for the decompositions in this chapter.

Recall that given the chain decomposition $\mathcal{C}$ with the chain cover property we can define a graph, the *chain cover graph*, $G(\mathcal{C}, \pi)$ whose vertices are the elements of the poset that $\mathcal{C}$ covers, and whose edges consist of the covering relations in the chains in $\mathcal{C}$ together with the *cover edges*, for each chain C, from $starter(C)$ to $\pi_s(C)$ and from $terminator(C)$ to $\pi'_t(C)$.

For a given n the constructions from Section 3.3 have one embedding and so no particular way of differentiating between different embeddings was required. We define the following to overcome this.

Definition. A *chain decomposition embedding* of a chain decomposition $\mathcal{C}$ with the chain cover property is a planar embedding of the chain cover graph $G(\mathcal{G}, \pi)$.

Such a chain decomposition embedding specifies the order of chains to be laid out in the plane (and thus the positioning of the cover edges); thus it makes sense, given such an embedding, to refer to the first chain C_1, second chain C_2, and so on, where the first chain is identified as one of the chains that is extremal along the axis the chains are laid out on. For example, Figure 5.2 shows a chain decomposition embedding of $\mathcal{B}_4$ with each chain embedded in a horizontal row and chains progressing vertically from C_1 uppermost to C_6 at the bottom. A chain embedded in a chain decomposition embedding can be referred to as a *row* of the chain decomposition embedding.

Without loss of generality, we will always index chains in an embedding from top to bottom or left to right (depending on the particular layout of the chains, whether horizontal or vertical).

5.3 Reverse Symmetric Chain Decomposition Embeddings

In this section we discuss a variation of the Greene Kleitman decomposition of the boolean lattice used in Section 3.3 to create monotone Venn diagrams. This variation proves to have a more interesting symmetry on the sphere. We define the construction recursively, as in [83].

Recall the notion of chain decomposition isomorphisms and dual-isomorphisms from the previous section. These can also be applied to chain decomposition embeddings.

Definition. Let E be a chain decomposition embedding of $\mathcal{B}_n$ with k chains, with chains numbered sequentially (as described in Section 5.2). Then E is *reverse symmetric* if the function $f(C_i) = C_{k+1-i}$ reversing the order of the chains, and applying $\phi(x) = \overline{x}$ to the elements is a dual-automorphism of the embedding.

Thus, the chain decomposition maps onto itself; *i.e* given a chain decomposition with m chains, the ith chain maps onto the $(m-i+1)$th chain in the decomposition and each element σ in the ith chain maps onto its complement $\overline{\sigma}$ in the $(m-i+1)$th chain. As an example of a reverse symmetric chain decomposition, in Figure 5.2, the first chain, $\{0100, 0101\}$, when complemented and reversed, gives $\{1010, 1011\}$, which is the last chain; the second chain, $\{0010, 0110\}$, when complemented and reversed, gives $\{1001, 1101\}$, the second last chain, and so on.

We modify the basis and recursive rule used to create the Greene Kleitman decomposition pattern to create a reverse symmetric chain decomposition embedding (RSCD), by proving that, as in the Greene Kleitman pattern, the RSCDs we construct have the chain cover property and planar embeddings of their chain cover graphs.

Basis. The basis for a recursive construction is a RSCD for $\mathcal{B}_2$, shown in Figure 5.1.

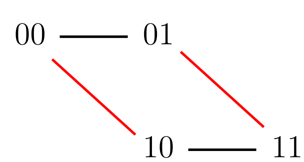


Figure 5.1: RSCD for $\mathcal{B}_2$, with chain cover edges shown in red

The chain cover edges in Figure 5.1 are the edges shown in red that connect different chains together; cover edges in the chains are shown in black.

The following recursive rules allow us to construct, given a RSCD for $\mathcal{B}_n$, a RSCD for $\mathcal{B}_{(n+1)}$.

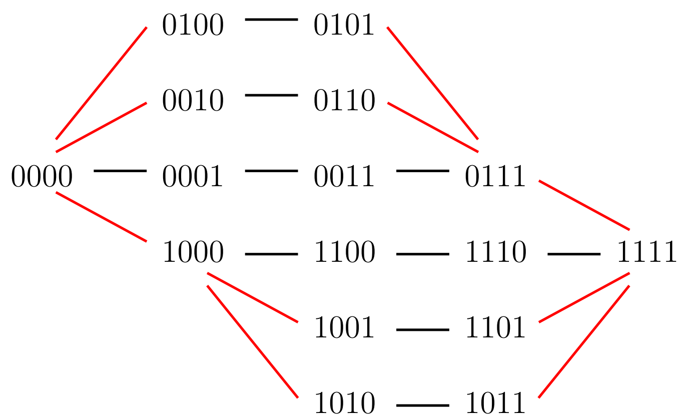


Figure 5.2: The RSCD for $\mathcal{B}_4$ generated by the recursive construction

Recursive Rule. Given a RSCD for $\mathcal{B}_n$ with m chains, the ith chain '$\sigma_1\ \sigma_2\ \ldots\ \sigma_k$', embedded as a row, is replaced in one of two ways, depending on whether the row is in the first half of the RSCD (*i.e.* whether $i \leq \frac{m}{2}$).

If $i \leq \frac{m}{2}$, the row becomes the two rows

$$\begin{array}{ccccc} & \sigma_2 0 & \ldots & \sigma_k 0 & \\ \sigma_1 0 & \sigma_1 1 & \ldots & \sigma_{k-1} 1 & \sigma_k 1, \end{array} \tag{5.1}$$

where if $k = 1$ the first row is omitted and the second becomes $\sigma_1 0\ \sigma_1 1$;

if $i > \frac{m}{2}$, the row becomes

$$\begin{array}{ccccc} \sigma_1 0 & \sigma_2 0 & \ldots & \sigma_k 0 & \sigma_k 1 \\ & \sigma_1 1 & \ldots & \sigma_{k-1} 1, & \end{array} \tag{5.2}$$

where if $k = 1$ the second row is omitted and the first becomes $\sigma_k 0\ \sigma_k 1$.

Knuth notes in [83] (when discussing the Greene-Kleitman construction), and thus we note also, that it is easy to show by induction that each of the 2^n bitstrings appears exactly once in the pattern, the bit strings with k 1s all appear in the same column, and within each row, consecutive bit strings differ by changing a 0 to a 1.

Each induction has two cases for a string in C_i, depending on if $i \leq \frac{m}{2}$, and note that a string in the first half remains in the first half after application of the recursive rule, and similarly for the second half. The first two properties imply that there are $\binom{n}{k}$ elements in each column.

Observation 2. *Given a RSCD M for $\mathcal{B}_n$ and applying the construction to give a RSCD M' for $\mathcal{B}_{(n+1)}$, chains in M become chains in M'.*

Proof. By definition of Rules 5.1 and 5.2. □

Lemma 15. *Given a RSCD M for $\mathcal{B}_n$ and applying the construction to give a RSCD M' for $\mathcal{B}_{(n+1)}$, if M has reverse symmetry, M' does also.*

Proof. Rules 5.1 and 5.2 locally replace rows, so since there is no global replacement it suffices to show that if a row C_i maps onto a row in M then the same occurs in M', and the global property of being reverse is preserved.

Given a row $\sigma = \sigma_1\ \sigma_2\ \ldots\ \sigma_k$ in the first $m/2$ rows of M, it maps onto its conjugate row $\omega = \omega_1\ \omega_2\ \ldots\ \omega_k$ in the second $m/2$ rows of M; *i.e.* $\overline{\sigma_1} = \omega_k$, $\overline{\sigma_2} = \omega_{k-1}$, etc. Applying Rule 5.1 to σ gives the row

$$\begin{array}{ccccc} & \sigma_2 0 & \ldots & \sigma_k 0 & \\ \sigma_1 0 & \sigma_1 1 & \ldots & \sigma_{k-1} 1 & \sigma_k 1. \end{array} \tag{5.3}$$

Since ω is in the second half of M Rule 5.2 is applied, which gives

$$\begin{array}{ccccc} \omega_1 0 & \omega_2 0 & \ldots & \omega_k 0 & \omega_k 1 \\ & \omega_1 1 & \ldots & \omega_{k-1} 1. & \end{array} \tag{5.4}$$

It is easy to verify that the chains in (5.3) are the complements of those in (5.4), and since their order is reversed all chains in the first half of M' will map onto their complementary chains in the second half of M', as desired. □

We can ignore the embedding to obtain the following lemma.

Lemma 16. *The chain decomposition of $\mathcal{B}_n$ given by the RSCD construction of Rules 5.1 and 5.2 is self-dual under the bijection $\phi : \phi(x) = \overline{x}$.*

Proof. By Lemma 15. ☐

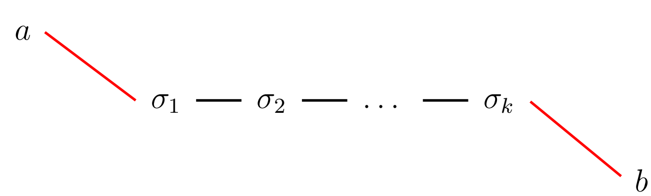
We now prove by induction that the RSCD for the n-cube has the chain cover property; the base case was shown in Figure 5.1.

Lemma 17. *Given a RSCD M for $\mathcal{B}_n$ and applying the construction to give a RSCD M' for $\mathcal{B}_{(n+1)}$, if M has the chain cover property, then the RSCD M' formed by applying Rules 5.1 and 5.2 also has the chain cover property.*

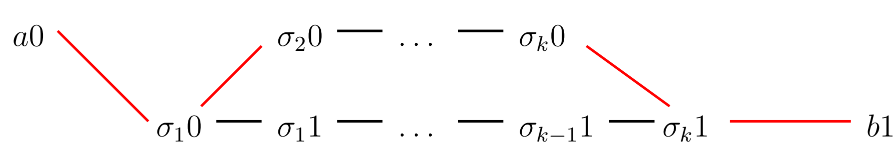
Proof. Since M has the chain cover property it suffices to show that the operation of replacing all rows according to Rules 5.1 and 5.2 creates new rows satisfying the chain cover property, and the chain cover edges will not introduce any edge crossings since the replacement rows and cover edges will be locally crossing-free.

Given a row $\sigma_1\ \sigma_2\ \ldots\ \sigma_k$ in M, there are nodes a and b such that a covers σ_1 and σ_k covers b (note that a and b do not have to be part of the same chain), and a and b are the starter and terminator of their respective chains. After applying rules 5.1 and 5.2, a becomes $a0$ at the start of a longer chain (regardless of which rule is applied), and b will become $b1$ at the end of a longer chain.

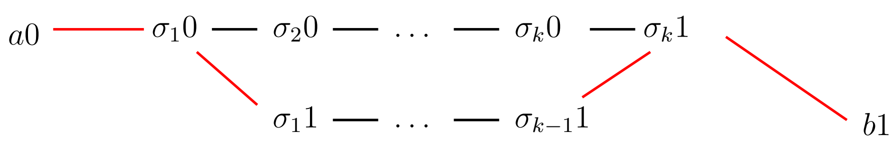
Given the chain



Rule 5.1 gives the chains:



and Rules 5.2 gives the chains:



If Rule 1 is applied, $a0$ covers $\sigma_1 0$ since a covered σ_1, $\sigma_k 1$ covers b_1 since σ_k covered b, $\sigma_1 0$ covers $\sigma_2 0$ since they differ by a single bit, and $\sigma_k 0$ covers $\sigma_k 1$ since they differ by a single bit.

If Rule 2 is applied, again $a0$ covers $\sigma_1 0$ and $\sigma_k 1$ covers b_1, $\sigma_1 0$ covers $\sigma_1 1$ since they differ by a single bit, and $\sigma_{k-1} 1$ covers $\sigma_k 1$ since they differ by a single bit.

None of the new edges introduce a crossing and so applying the expansion to all rows in M gives a RSCD M' with the chain cover property. $\square$

Theorem 6. *There exists a reverse symmetric chain decomposition embedding of $\mathcal{B}_n$, for all $n \geq 2$, such that*

- *the number of chains is $N(n)$, as calculated in the next section in Theorem 7, and*

- *the chain decomposition has the chain cover property.*

Proof. The basis in Figure 5.1 is reverse symmetric and has the chain cover property. Then build the construction as in Rules 5.1 and 5.2, and repeatedly apply Observation 2 and Lemmas 15 and 17, with the induction over n. $\square$

In general there are many RSCDs for a given n; this construction, starting from the base case shown for $n = 2$, gives us just one of them (another diagram is shown in Section 5.6).

5.3.1 Counting Results

In this section we count the number of chains in the RSCD construction given in the previous section. The Greene Kleitman decomposition, it is easy to prove, has the number of chains equal to the middle binomial coefficient $\binom{n}{\lfloor n/2 \rfloor}$ (see [83]).

Let $N(n, k)$ be the number of chains in the construction of length k. Also define the total number of chains in the construction to be $N(n) = \Sigma_{1 \le k \le n} N(n, k)$. Directly from the recursive construction, we have the recurrence relation

$$N(n,k) = \begin{cases} 2 & \text{if } n = k \geq 1 \\ N(n-1,2) & \text{if } k = 1 \\ N(n-1,k-1) + N(n-1,k+1) & \text{if } k > 1 \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 18. *If n and k have the same parity, then*

$$N(n,k) = 2\frac{k}{n}\binom{n}{(n-k)/2}$$

and, otherwise, $N(n, k) = 0$.

Proof. This recursive formula is the same as that for the Catalan triangle [119, Sequence A053121], with double the base case. The reference [119, Sequence A053121] gives the closed form

$$N(n,k) = \frac{k}{n}\binom{n}{(n-k)/2},$$

for the Catalan triangle, and twice this is the formula we want. □

Corollary 5. $N(n, 1) = 2\mathbf{C}_{(n-1)/2}$, *twice the $(n-1)/2$th Catalan number, if n is odd, otherwise it is 0.*

Proof. If n is odd, let $m = 2n+1$. Then

$$\begin{aligned} N(n,1) &= \frac{2}{n}\binom{n}{(n-1)/2} \qquad \text{by the previous lemma,} \\ &= 2\frac{1}{2m+1}\binom{2m+1}{m} \\ &= 2\frac{1}{m+1}\binom{2m}{m}, \end{aligned}$$

which is twice a well-known expression for the Catalan number $\mathbf{C}_m$. If n is even, the parity of n and k differ, so by the previous lemma $N(n,1) = 0$. □

Lemma 19. *If n is even, all chains in the chain decomposition given by the RSCD construction for n have even length, otherwise all chains have odd length.*

Proof. By induction. Figure 5.1 shows it is true for $n = 2$, and it can be easily observed to be true for $n = 3$ by applying the construction. Assume the lemma is true for $n-1$ and $n-2$. If n is even, assume the construction for $n-1$ has all odd-length chains. The replacement rule replaces a chain of length k, with k odd, with two chains of length $k+1$ and $k-1$ respectively, so their lengths must both be even, or one chain of length two if $k = 1$. In both cases the resulting chains all have even length. The case for n odd is similar. □

Theorem 7.

$$N(n) = \begin{cases} \binom{n}{n/2} & \text{if } n \text{ even, } n > 1 \\ 2\binom{n-1}{(n-1)/2} & \text{if } n \text{ odd, } n > 2 \end{cases}$$

Proof. By induction. The basis for $n = 2$ is shown in Figure 5.1 and the construction for $n = 3$ is easily generated. We consider the two cases, n odd and even.

For n odd, the number of chains in the construction for $n-1$ is $\binom{n-1}{(n-1)/2}$ as $n-1$ is even. By Lemma 19 all chains in the construction for $n-1$ are of even length and so are all replaced by two odd chains; from the construction it is apparent that each even-length chain gives exactly two odd-length chains. Thus for n odd, $N = 2\binom{n-1}{(n-1)/2}$.

For n even, the construction for $n-1$ gives $2\binom{n-2}{(n-2)/2}$ chains by induction, and all chains are odd length. Applying the construction rule gives two chains for every chain in the construction for $n-1$, except for chains of length one which will give one chain. By Lemma 5 the number of length one chains is

$$N(n-1,1) \;=\; 2\mathbf{C}_{\frac{n-2}{2}} \;=\; \frac{2}{(n-2)/2+1}\binom{n-2}{(n-2)/2} \;=\; \frac{4}{n}\binom{n-2}{(n-2)/2}\,.$$

Thus the number of chains for n even is

$$\begin{aligned} N(n) &= 2N(n-1) - N(n-1,1) \\ &= 2\cdot 2\binom{n-2}{(n-2)/2} - \frac{4}{n}\binom{n-2}{(n-2)/2} \\ &= 4(\frac{n-1}{n})\binom{n-2}{(n-2)/2}, \text{ which is easily reduced to} \\ &= \binom{n}{n/2}, \text{ as desired.} \end{aligned}$$

□

Recall that the minimum number chains possible in a chain decomposition is $\binom{n}{\lfloor n/2 \rfloor}$ (see Section 3.3.2). The previous results have shown that the construction may, depending on n, give a chain decomposition with more than the minimum number of chains possible.

Corollary 6.

$$N(n) - \binom{n}{\lfloor n/2 \rfloor} \;=\; \mathbf{C}_{\frac{n-1}{2}} \;\textit{if } n \textit{ odd, } n > 2\ ,$$

that is, when n is odd, the difference between the number of chains produced by the PSCD construction and the minimum number of chains in a chain decomposition of

$\mathcal{B}_n$ is the $(n-1)/2$th Catalan number.

Proof. If n is odd,

$$N(n) - \binom{n}{\lfloor n/2 \rfloor} \;=\; 2\binom{n-1}{(n-1)/2} - \binom{n}{(n-1)/2}$$

which is easily reduced to $\mathbf{C}_{\frac{n-1}{2}}$. □

5.4 Antipodally Symmetric Chain Decomposition Embeddings

In this section we show how the construction for decompositions given in the previous section can be easily modified to give a chain decomposition that has symmetry under an antipodal map, which we refer to as an *antipodally symmetric chain decomposition embedding* (ASCD).

Definition. Let E be a chain decomposition embedding of $\mathcal{B}_n$ with k chains, k an even number, with chains numbered sequentially (as described in Section 5.2). Then E is *antipodally symmetric* if the function $f(C_i) = C_{i+k/2 \mod k}$ swapping the first and second halves of the chains, and applying $\phi(x) = \overline{x}$ to the elements is a dual-automorphism of the embedding.

Basis. The basis for a recursive construction is an ASCD for $\mathcal{B}_2$, shown in Figure 5.1 in the previous section.

The modification of the earlier construction simply consists of reordering the rows in the second half of the embedding; the chains in the ASCD construction are the same as in the RSCD construction.

Similar to the rules in the previous section, the following recursive rules allow us to construct, given an ASCD for the n-cube $\mathcal{B}_n$, an ASCD for the $(n+1)$-cube $\mathcal{B}_{(n+1)}$.

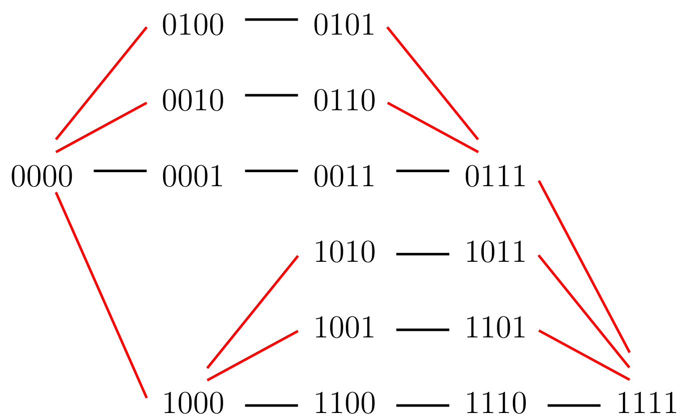


Figure 5.3: Chain decomposition embedding symmetric under antipodal map, generated by the recursive construction, for $\mathcal{B}_4$; chain cover edges are shown in red

Recursive Rule. Given an ASCD for $\mathcal{B}_n$ with m rows, the ith row '$\sigma_1\ \sigma_2\ \ldots\ \sigma_k$' is replaced in one of two ways, depending on whether the row is in the first half of the ASCD (*i.e.* whether $i \leq \frac{m}{2}$).

If $i \leq \frac{m}{2}$, the row becomes the two rows

$$\begin{array}{ccccc} & \sigma_2 0 & \ldots & \sigma_k 0 & \\ \sigma_1 0 & \sigma_1 1 & \ldots & \sigma_{k-1} 1 & \sigma_k 1, \end{array} \tag{5.5}$$

where if $k = 1$ the first row is omitted and the second becomes $\sigma_1 0\ \sigma_1 1$; this is identical to the first rule used in the recursive construction for RSCDs in the previous section.

If $i > \frac{m}{2}$, the row becomes

$$\begin{array}{ccccc} & \sigma_1 1 & \ldots & \sigma_{k-1} 1, & \\ \sigma_1 0 & \sigma_2 0 & \ldots & \sigma_k 0 & \sigma_k 1 \end{array} \tag{5.6}$$

where if $k = 1$ the first row is omitted and the second becomes $\sigma_k 0\ \sigma_k 1$. This is similar to the second rule for RSCDs in the previous section, except the order of the rows is reversed.

It is easy to prove, in the same way as the previous section, that the decomposition has all the right properties.

Observation 3. *Given an ASCD M for $\mathcal{B}_n$ and applying the construction to give an ASCD M' for $\mathcal{B}_{(n+1)}$, chains embedded as rows in M become chains embedded as rows in M'.*

Proof. By definition of Rules 5.5 and 5.6. □

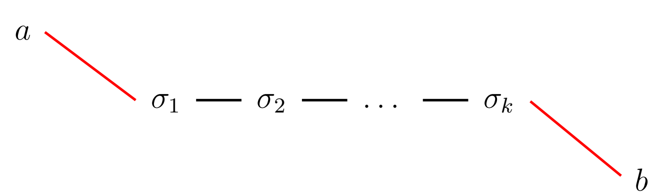
Lemma 20. *Given an ASCD M for $\mathcal{B}_n$ and applying Rules 5.5 and 5.6 to give an ASCD M' for $\mathcal{B}_{(n+1)}$, if M has antipodal symmetry, M' does also.*

Proof. The proof is almost identical to that of Lemma 15, except that the order of the rows in Equation 5.4 switch. □

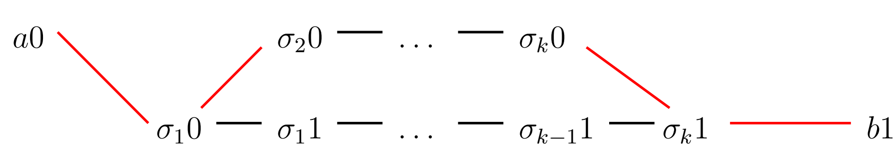
Lemma 21. *Given an ASCD M for $\mathcal{B}_n$ and applying Rules 5.5 and 5.6 to give an ASCD M' for $\mathcal{B}_{(n+1)}$, if M has the chain cover property, then M' also has the chain cover property.*

Proof. The proof is almost identical to that of Lemma 17 above. The chain cover edges shown in the proof of Lemma 17 change to reorder the chains in the application of Rule 5.6.

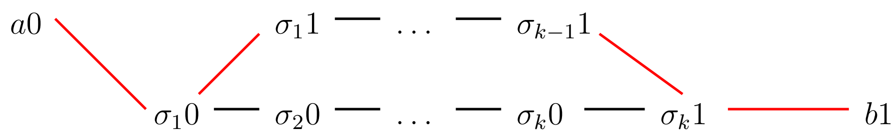
Given the chain



Rule 5.5 gives the chains:



and Rule 5.6 gives the chains:



□

Theorem 8. *There exists a chain decomposition embedding of $\mathcal{B}_n$ that is symmetric under an antipodal map, for all $n \geq 2$, such that*

- *the number of chains is $N(n)$ (see Theorem 7),*
- *the chain decomposition embedding has the chain cover property.*

Proof. The basis in Figure 5.1 is antipodally symmetric and has the chain cover property. Then build the construction by inductively applying Rules 5.5 and 5.6, and repeatedly apply Observation 3 and Lemmas 20 and 21, with induction over n. □

All of the counting results from Section 5.3.1 still apply, since the only difference in the construction is to reorder half of the chains in the embedding.

5.5 Diagrams from Chain Decomposition Embeddings

In this section we discuss the Venn diagrams that are created on the sphere by taking the dual diagrams of the chain decomposition embeddings discussed in Sections 5.4 and 5.3. These constructions give diagrams with interesting total symmetries (*i.e* all curves map onto themselves).

Recall Lemma 3 from Section 3.3:

Lemma 3 ([54, p. 8])**.** *Let $\mathcal{C}$ be a symmetric chain decomposition with the chain cover property for $\mathcal{B}_n$, let π be a chain cover mapping for $\mathcal{C}$, and let $P(\mathcal{C}, \pi)$ be the planar embedding of $G(\mathcal{C}, \pi)$ described in the proof of Lemma 2. Then the geometric* *dual of $P(\mathcal{C},\pi)$, called $P^*(\mathcal{C},\pi)$, is a monotone Venn diagram of n curves with the minimum number of vertices.*

This result can easily be changed to accommodate chain decompositions embedded on the surface of the sphere, as follows in subsequent sections. First, we enhance the definition of the embeddings as follows:

Definition. A *spherical chain decomposition embedding* of a chain decomposition $\mathcal{C}$ with the chain cover property is an embedding $P(\mathcal{C},\pi)$ of the chain cover graph $G(\mathcal{C},\pi)$ on the 2-sphere.

It is clear that any graph G with a planar embedding also has an embedding on the sphere, provided, if a cylindrical projection is used, the planar embedding of G does not map more than one point onto each line which is an image of a pole on the sphere. Any planar embedding of G can be projected onto the sphere by a reverse stereographic projection to obtain a spherical embedding. We will use reverse cylindrical projections in the following sections to produce particular spherical chain decomposition embeddings of the chain decomposition embeddings described in the previous sections, and then show that the diagrams produced have some symmetries not present on the plane.

5.5.1 Venn Diagrams with Rotational Symmetry on the Sphere

In this section we discuss the diagrams created as the duals of the reverse symmetric chain decomposition embeddings discussed in Section 5.3. Note that for these diagrams we depart from our normal convention that the singular axis of rotation passes through the north and south poles of the sphere (with polar coordinates $\phi = \pi/2$ and $-\pi/2$ respectively), since the cylindrical projection we discuss is more intuitive using the coordinates and axis of rotation as specified.

The following proof establishes the resulting symmetry from the n-Venn diagrams on the sphere created from the construction of the previous section.

Lemma 22. *Given n, let $\mathcal{C}$ be a RSCD for $\mathcal{B}_n$ with the chain cover mapping π and the planar embedding $P(\mathcal{C}, \pi)$. Then there exists a vector $\vec{v}$ such that $P(\mathcal{C}, \pi)$ has an embedding on the 2-sphere such that the geometric dual of $P(\mathcal{C}, \pi)$, called $P^*(\mathcal{C}, \pi)$, is a monotone Venn diagram of n curves with total symmetry group $S_T(P^*(\mathcal{C}, \pi)) = \mathsf{C}_2$ under the rotation $R_{\vec{v},\pi}$.*

Proof. We first describe the embedding of $\mathcal{C}$ onto the plane, its labelled, dual graph, and then specify the projection onto the sphere giving the final diagram. On the plane, the ith chain $C_i \in \mathcal{C}$ is embedded vertically (with all vertices on a line $x = i$) such that a vertex x on C has y-coordinate $-rank(x)$ and x-coordinate i. All edges of $\mathcal{C}$ are embedded as straight lines. The chains and chain cover edges are thus ordered horizontally as in the construction of Section 5.3, from left to right in increasing order of i, with all nodes of equal rank at the same horizontal level, and edges embedded vertically; the chain cover edges are embedded as straight lines between nodes differing in rank by one. Note that this embedding is essentially the same as that of the GKS construction [54], except with a lack of mirror symmetry between the chain cover edges. The line $y = -rank(C)/2$, *i.e* the horizontal line through the elements of rank $n/2$ (if n is even) or halfway between the elements of rank $(n-1)/2$ and $(n+1)/2$ (if n is odd), is referred to as the *equator* of the embedding.

We next construct the geometric dual $P^*(\mathcal{C}, \pi))$ of $\mathcal{C}$, noting that we overload the notation P^* to refer to the planar dual of $P(\mathcal{C}, \pi)$ as well as the dual of the embedding on the sphere of $P(\mathcal{C}, \pi)$; it will be clear from context which surface embedding is being referred to. The dual on the plane is also constructed as in the GKS construction. With one exception, discussed in the following paragraph, vertex f^* of P^* is placed in the interior of face f of P, such that the x-coordinate of f is $i + 1/2$, where C_i and C_{i+1} are the two chains embedded on either side of f, and the y-coordinate of f^* is such that it lies on the equator. We can refer to the vertices $f^* \in P^*$ that are embedded in the faces of P on the equator as *face vertices*.

Let $m = N(n)$ be the number of chains. Since the chains are spaced equally according to their index by the construction, the leftmost, or first, chain, with minimum x-coordinate, is C_1 and the rightmost, or last, is C_m. Let t refer to the external face on the plane; embed the vertex $t^* \in P^*$ at the point on the equator with x-coordinate $1/2$, and also embed an image of t^* on the equator at x-coordinate $m + 1/2$. These two images will be identified under the reverse cylindrical projection onto the sphere that will be performed on P^* applied to the rectangle bounded to either side by the lines $x = 1/2$ and $x = m + 1/2$.

Example. Figure 5.4(a) shows the construction of P^*, with edges drawn as smooth curves for clarity. The images of t^* are identified as v' in Figure 5.4.

To construct the edges of $P^*(\mathcal{C}, \pi)$, a point is chosen on each edge of P where the edge of P^* will cross it. For each face f of P and each edge e on on its boundary, a half-edge of P^* is drawn from f^* to the crossing point so that the half-edges incident with f^* are internally disjoint. The two half-edges of P^* meet at the crossing point and together form the edge $e^* \in P^*$, and e^* is labelled with the index x of the bit that changes between the endpoints of e, so $e = (S, S \cup \{x\})$. Edges incident to the extremal point t^* are drawn such that edges crossing chain C_1 and its cover edges are incident to the image of t^* at $x = 1/2$, and edges crossing C_m and its cover edges are incident to the image of t^* at $x = m + 1/2$;

Now construct the reverse cylindrical projection onto the sphere of $P^*(\mathcal{C}, \pi)$ about a rectangle bounded by $x = 1/2$, $x = m + 1/2$, $y = 0$, and $y = -n$; define the projection by $\theta = \frac{2\pi(x-1/2)}{m}$ and $\phi = \pi y/n + \pi/2$. This projection maps the two images of t^* onto the point at $(\theta, \phi) = (0, 0)$ and the equator maps to the sphere's equatorial line $\phi = 0$. The face that is the dual of the vertex of P corresponding to the empty set, labelled 0^n, includes the north pole at $\phi = \pi/2$, and the face corresponding to the complete set, labelled 1^n, contains the south pole at $\phi = -\pi/2$.

Using the proof of GKS [54] it is straightforward to show that $P^*(\mathcal{C}, \pi)$ on the sphere has the properties required of an n-Venn diagram; to show that the edges labelled i in P^* form a simple cycle and thus form a simple closed curve the technique is to show that the corresponding edges of P on the sphere form a *bond*, that is, a minimal set of edges whose removal disconnects P. Furthermore P^* is an n-Venn diagram on the sphere, which can be shown by considering the 2^n vertices of P that correspond to the 2^n faces of P^*. Also note that in P each vertex of $rank(k)$, $0 < k < n$, is adjacent to a vertex of rank $k-1$ and a vertex of rank $k+1$, which implies that in P^* each region of weight k, where $0 < k < n$, is adjacent to a region of weight $k-1$ and a region of weight $k+1$, and thus $P^*(\mathcal{C}, \pi)$ is a monotone n-Venn diagram.

We now consider the symmetries of P^* on the sphere. Define the vector $\vec{v}$ from the centre of the sphere through the point $v = (\theta = \pi, \phi = 0)$ on the equator and antipodal to the image v' of t^* at $(0,0)$; this vector defines an axis of rotation through the points v, v'. This axis passes through the face vertex $f^* \in P^*$ that is embedded in the face between the two chains $C_{m/2}$ and $C_{m/2+1}$, and, as noted, through the vertex $v' \in P^*$ between the chains C_1 and C_m. The rotational symmetry of period two about this axis is $R_{\vec{v},\pi}$. The action of $R_{\vec{v},\pi}$ on a point $p = (\theta, \phi)$ is $R_{\vec{v},\pi}(p) = (2\pi - \theta, -\phi)$.

Now consider a non-terminator vertex $p \in C_i$ with label $p = p_1 p_2 \dots p_n$ corresponding to a set S; it is embedded on the planar embedding P at position $(x, y) = (i, -rank(p))$, and the embedding on the 2-sphere puts p at position

$$(\theta, \phi) = (\tfrac{2\pi(i-1/2)}{m}, \tfrac{-\pi rank(p)}{n} + \pi/2).$$

Consider a neighbouring vertex to p, called q, on the same chain C_i, with $rank(q) = rank(p) + 1$ and labelled corresponding to set $S \cup \{x\}$; on the sphere q is at position

$$(\theta, \phi) = (\tfrac{2\pi(i-1/2)}{m)}, \tfrac{-\pi(rank(p)+1)}{n} + \pi/2).$$

The edge $e = (p, q)$ is embedded as an arc on the line of longitude $\theta = \frac{2\pi(i-1/2)}{m}$, and the edge e^* of P^*, with $label(e^*) = x$, crosses e at the chosen point (the image under the projection of the crossing point chosen to build P^* on the plane).

Under the operation of $R_{\vec{v},\pi}$, the points p and q are taken to

$$\begin{aligned} R_{\vec{v},\pi}(p) &= (2\pi - \frac{2\pi(i-1/2)}{m}, -(\frac{-\pi rank(p)}{n} + \pi/2)) \\ &= (\frac{2\pi(m-i+1/2)}{m}, \frac{\pi rank(p)}{n} - \pi/2) \end{aligned}$$

and

$$R_{\vec{v},\pi}(q) = (\frac{2\pi(m-i+1/2)}{m}, \frac{\pi(rank(p)+1}{n} - \pi/2) \ .$$

Now consider the reverse symmetry of the embedding P, in which the operation of reversing the chains, reversing the order of the chains, and complementing the elements of the chain is an automorphism of the embedding. Thus, the vertex $p \in C_i$ under the reverse symmetry on the chain decomposition embedding maps to the point p' on the chain C_{m-i+1} with $rank(p') = n - rank(p)$, and the set corresponding to the label of p' is $\overline{S}$. The position of p' in P^* on the plane is $(x, y) = (m-i+1, -(n - rank(p))) = (m-i+1, rank(p)-n)$. Now, under the cylindrical projection onto the 2-sphere, p' is at position

$$\begin{aligned} (\theta, \phi) &= (\frac{2\pi(m-i+1-1/2)}{m)}, \frac{\pi(rank(p)-n)}{n} + \pi/2) \\ &= (\frac{2\pi(m-i+1/2)}{m)}, -\pi + \frac{\pi(rank(p))}{n} + \pi/2) \\ &= (\frac{2\pi(m-i+1/2)}{m)}, \frac{\pi(rank(p))}{n} - \pi/2) \ . \end{aligned}$$

Similarly, the image of q, called q', projects to

$$(\theta, \phi) = (\frac{2\pi(m-i+1/2)}{m)}, \frac{\pi(rank(p)+1)}{n} - \pi/2) \ ,$$

and so $R_{\vec{v},\pi}(p) = p' \in P^*$ and $R_{\vec{v},\pi}(q) = q' \in P^*$.

A similar argument holds if (p, q) is a chain cover edge, the only difference in the above is that the y-coordinate for $q \in P$ is not the same as that of p, so the

ϕ-coordinate of $q \in P$ on the sphere is different than for p.

The points chosen for the crossing points of the edges in P^* have the same symmetry as they are chosen consistently in the chain edges and chain cover edges throughout P. Recall also the face vertices f^* chosen on the equator as vertices in P^*, where f^* has x-coordinate $\frac{i+(i+1)}{2}$, where C_i and C_{i+1} are the chains on either side of f^*. These points are the endpoints of the edges $e^* \in P^*$, and since on the sphere the point f^* is located at coordinates $(\theta, \phi) = (\frac{2\pi(2i+1/2)}{2m}, 0)$, then $R_{\vec{v},\pi}(f^*) = g^*$, where g^* is some similar face vertex ($g^* = f^*$ itself if $i = m/2$, otherwise it is a different face vertex).

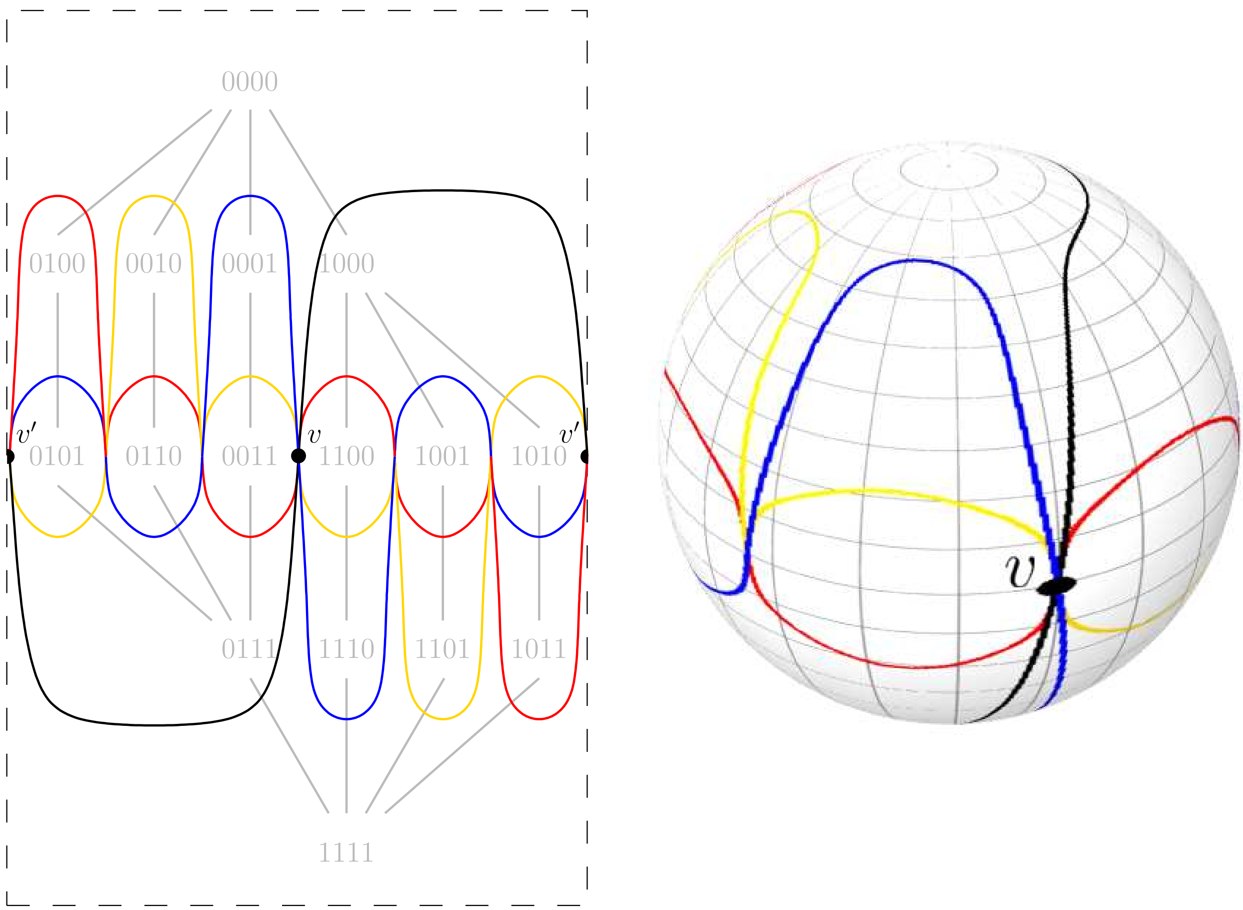
Now consider the edge $e^* \in P^*$ with $label(e^*) = x$ such that p has the label corresponding to subset S and q has the label corresponding to subset $S \cup \{x\}$.

The edge $e^{*'} \in P^*$ on the sphere that is the image of e^* under $R_{\vec{v},\pi}$ has some label, call it y. Let $label(R(p)) = T$ and $label(R(q)) = T - \{y\}$, for some subset T. By the definition of the chain decomposition embedding, $label(R(p)) = \overline{S}$ and $label(R(q)) = \overline{S \cup \{x\}}$. Let $T = \overline{S}$; then $label(R(q)) = \overline{S \cup \{x\}} = T - \{x\}$, and so $y = x$. In other words, the bit that changes over the edge $(R(p), R(q)) = e^{*'} \in P^*$ is x, and so $label(e^{*'}) = label(e^*) = x$.

We have established that the image of endpoints of e^*, which are the face vertices, are face vertices also, and the point where e^* crosses the edge (p, q) is preserved in the edge $(R_{\vec{v},\pi}(p), R_{\vec{v},\pi}(q))$, and if the edges are drawn consistently throughout the embedding then the arcs joining these points will be preserved also, and finally we have established that the label of the image $e^{*'}$ of e^*, which is the colour of the edge, is the same as e^*. This establishes that the image of an edge of P^* under $R_{\vec{v},\pi}$ is another edge of the same colour (for example, see Figure 5.4(a)). This establishes that $R_{\vec{v},\pi}$ is a total symmetry for P^*. $\square$

Figure 5.4 shows the diagram constructed from the RSCD construction from Section 5.3. The axis of rotation for the rotational symmetry passes through the vertices

v and v' on opposite sides of the sphere at the equator.



(a) Cylindrical projection of diagram

(b) Diagram on sphere

Figure 5.4: 4-Venn diagram with rotational symmetry C_2 from the RSCD for $\mathcal{B}_4$, with axis of rotation passing through vertices v and v'

5.5.2 Antipodally Symmetric Venn Diagrams

In this section we discuss the diagrams created as the dual of the antipodally symmetric chain decomposition embeddings discussed in Section 5.4.

The following proof establishes the resulting antipodal symmetry from the n-Venn diagrams on the sphere created from the construction of the previous section. An interesting situation that does not occur in the proof of Lemma 22 is that if the two adjacent vertices p and q in the antipodally symmetric chain decomposition embedding are such that (p, q) is a chain cover edge and $label(p) = 0^n$, then q may be such that the edge (p,q), under the glide reflection $G_{\vec{v},\pi}((p,q))$ on the sphere (as we discuss in the proof), crosses the line of longitude at $\theta = 0$. Representing this edge on the embedding of P^* in the plane is awkward, and we have always chosen to break the symmetry in the planar representation to preserve continuity of this chain cover edge by drawing it in a continuous fashion—meaning that if the cylindrical projection is taken of P^* on the sphere as we have described it *back* onto the plane, the edge $(G_{\vec{v},\pi}(p), G_{\vec{v},\pi}(q))$ would cross the edge of the enclosing rectangle, and so we have always chosen to draw it to be contained in that rectangle. For example, in Figure 5.5(a) the edge $(0111, 1111)$ is drawn contiguously within the inscribing rectangle instead of angling downwards to the left from 0111, across the rectangle boundary, to 1111. Note again that this breaking of symmetry does not apply to the spherical embedding of P^* and so it does not affect the following proof.

Lemma 23. *Given n, let $\mathcal{C}$ be an ASCD for $\mathcal{B}_n$ with the chain cover mapping π and the planar embedding $P(\mathcal{C},\pi)$. Then $P(\mathcal{C},\pi)$ has an embedding on the 2-sphere such that the geometric dual of $P(\mathcal{C},\pi)$, called $P^*(\mathcal{C},\pi)$, is a monotone Venn diagram of n curves with monodromic total symmetry group $S_T(P^*(\mathcal{C},\pi)) = \mathsf{S}_2$ under the glide reflection $G_{\vec{v},\pi}$ (also called inversion symmetry), for any vector $\vec{v}$.*

Proof. This proof is almost identical to the proof of Lemma 22, since the constructions and resulting diagrams are very similar.

The construction on the plane and reverse cylindrical projection are the same, taking into account the differences in the chain layout for the antipodal symmetric chain decomposition embedding, and the resulting diagram is a monotone n-Venn diagram on the sphere.

The key difference for the remainder of the proof is that to define the symmetry $G_{\vec{v},\pi}$, any vector v can be used, but for convenience we can choose the polar axis as is our usual convention, such that $\vec{v}$ is defined from the centre of the sphere through the point $\phi = \pi/2$ at the north pole. Thus the action of $G_{\vec{v},\pi}$ on a point $p = (\theta, \phi)$

is $G_{\vec{v},\pi}(p) = (\theta + \pi, -\phi)$. Defining the vertices p and q as in the proof of Lemma 22, it is easy to verify that the image of p under the antipodal symmetry of the chain decomposition embedding, and the image p' of p under $G_{\vec{v},\pi}$, is the point

$$G_{\vec{v},\pi}(p) = (\frac{2\pi(i-1/2)}{m} + \pi, \frac{\pi rank(p)}{n} - \pi/2)) \ ,$$

with a similar equation for q.

The same considerations of the images of the points in P^* hold for the edge $e^* \in P^*$ between p and q, and the same arguments show that the endpoints of $e^{*'}$ are preserved, the crossing point is preserved, the arcs are preserved, and $label(e^*) = label(e^{*'})$, and thus $G_{\vec{v},\pi}$ is a total symmetry for P^*, and so the resulting diagram has monodromic total symmetry group $S_T(P^*) = \mathsf{S}_1$. □

Figure 5.5 shows the diagram constructed from the ASCD construction from Section 5.4.

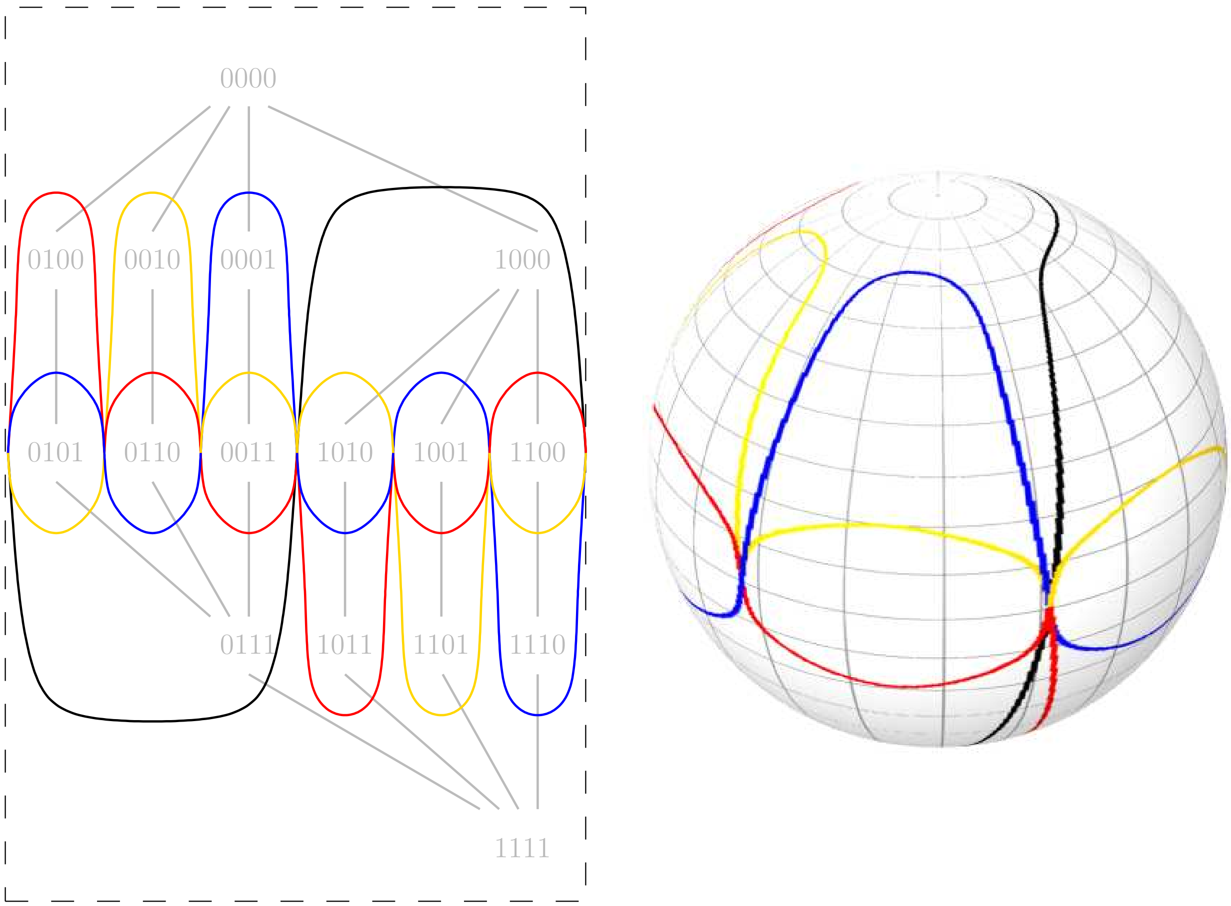
5.6 Other Diagrams

The diagram in Figure 5.6 realizes symmetry group C_2 under the antipodal symmetry operation, similar to the diagrams in Section 5.4. This diagram illustrates that a chain may be covered by an element from one chain and cover an element from another, as there is nothing in the definition of chain cover property that prevents this (note that this differs from the usage in [54]).

Figure 5.7 shows the dual graph of the diagram in Figure 5.6: compare, for example, with Figure 5.3.

5.6.1 Open Questions

We noted in Section 5.3.1 that the construction given for reverse symmetric and antipodally symmetric chain decompositions gives more than the minimum number of chains, starting with $n \geq 5$. The chain decomposition embedding with antipodal



(a) Cylindrical projection of diagram

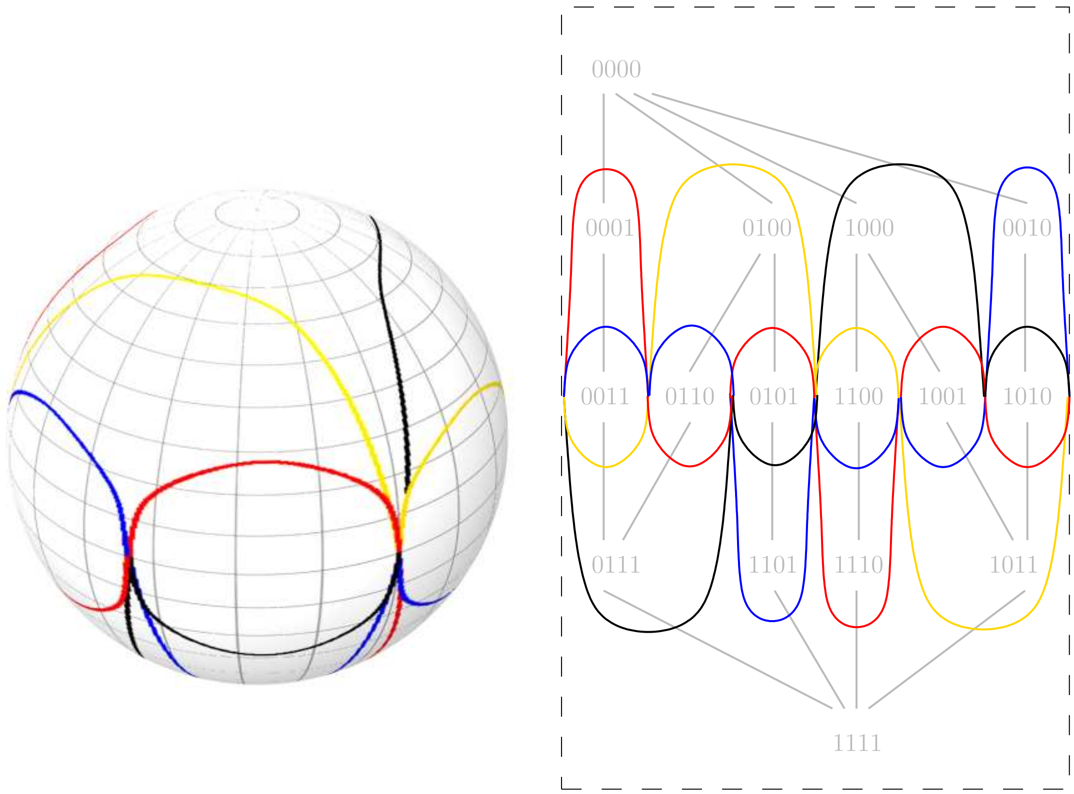
(b) Diagram on sphere

Figure 5.5: Antipodally symmetric 4-Venn diagram from the ASCD for $\mathcal{B}_4$

symmetry shown in Figure 5.8 for $n = 5$ contains $\binom{5}{\lfloor 5/2 \rfloor} = 10$ chains, as opposed to the number $N(5) = 12$, from Section 5.3.1, given by the construction. Of course, a reverse symmetric chain decomposition can be easily created from that in Figure 5.8 by reversing the order of the second half of the chains. Since this construction has the chain cover property, it then gives another 5-Venn diagram with the symmetries described in Lemmas 22 and 23.

A natural question to ask is thus, do there exist reverse symmetric (or antipodally symmetric) chain decompositions for any n odd with $\binom{n}{(n-1)/2}$ chains? If so, a construction similar to that for RSCDs (or ASCDs) would be interesting.

Furthermore it is easy to see that the resulting diagrams from our construction, as



(a) Diagram on the sphere (b) Cylindrical projection of the diagram

Figure 5.6: Different 4-Venn diagram with antipodal symmetry

developed in Section 5.5, are not simple, as usually the centre point between chains $C_{m/2}$ and $C_{m/2+1}$ (the point of rotation in the reverse symmetric diagrams) is a vertex of degree $2n$ in the resulting n-Venn diagram. Simple diagrams of this type seem to be hard to find, and it would be interesting to investigate when, or if, they exist.

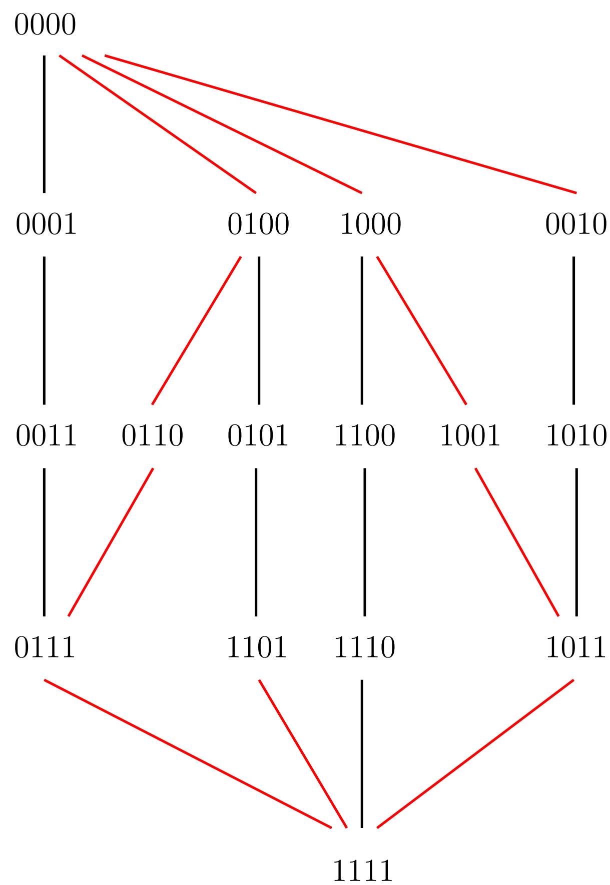


Figure 5.7: Chain construction of dual of 4-Venn diagram in Figure 5.6 with antipodal symmetry

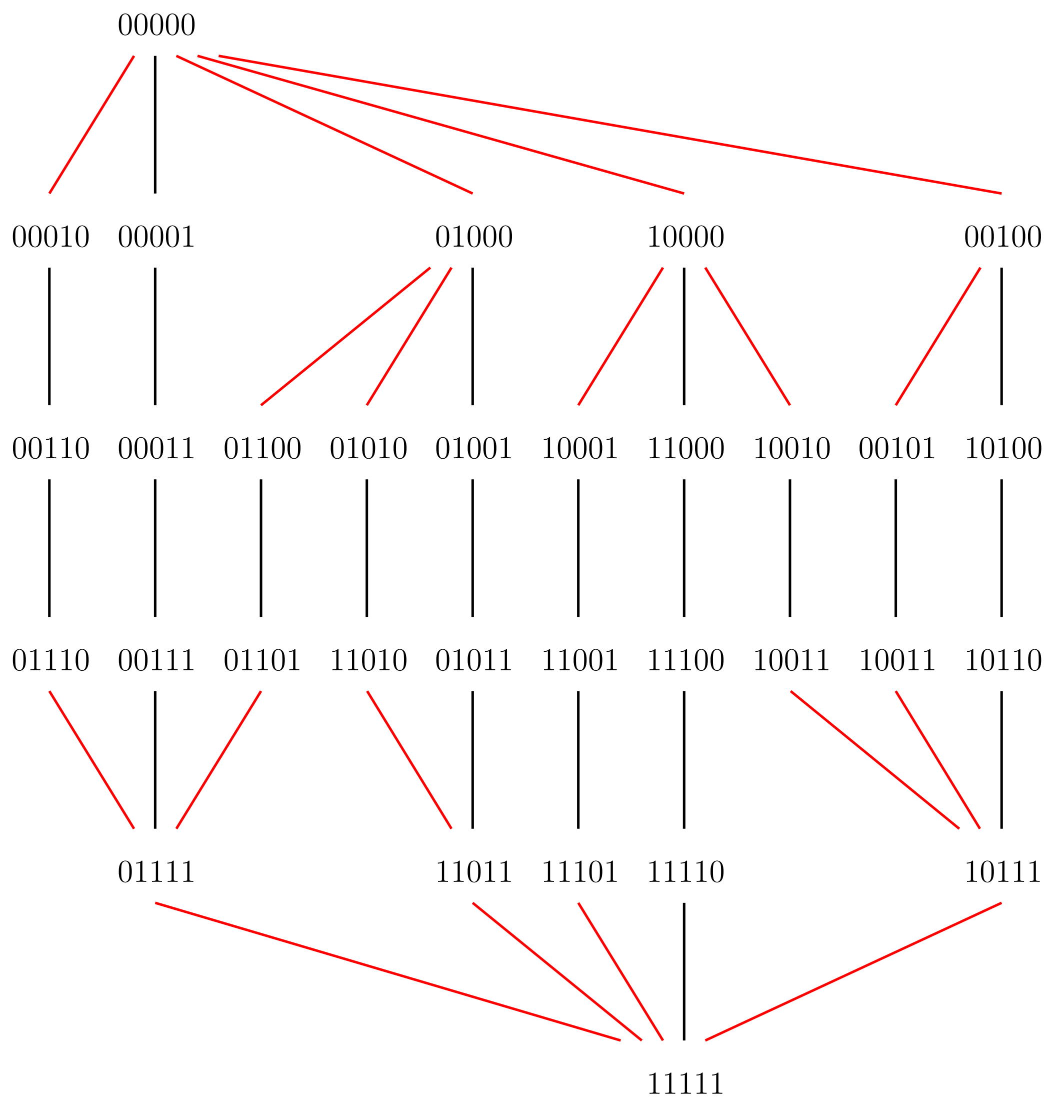


Figure 5.8: ASCD of $\mathcal{B}_5$ with minimum number of chains

Chapter 6

Total Symmetric Diagrams in Higher Dimensions

In this chapter we explore some diagrams in higher dimensions that have interesting total symmetry groups. The main result in this chapter is a construction, given any symmetry group that is a subgroup of a specific direct product group, of a higher-dimensional Venn diagram that realizes that symmetry group.

6.1 History of Diagrams in Higher Dimensions

In this section we give an overview of research in higher-dimensional diagrams, especially Venn diagrams and independent families.

A first approach is generalizing the familiar 3-circle diagram of Figure 2.3 to higher dimensions. A simple counting argument based on the fact that two different circles can intersect at most twice tells us that four distinct circles cannot be used to create a Venn diagram in two dimensions, as at most 14 separate regions can be created (this has been known since Venn [126], see for example [56]). However, in three dimensions, four overlapping 2-spheres[1] can form a 4-Venn diagram [93], by arranging them with their centres situated at the points of a regular tetrahedron, as

[1]This configuration of spheres occurs often in nature, most commonly in sand as a silica (SiO_3) molecule.

in Figure 6.1.

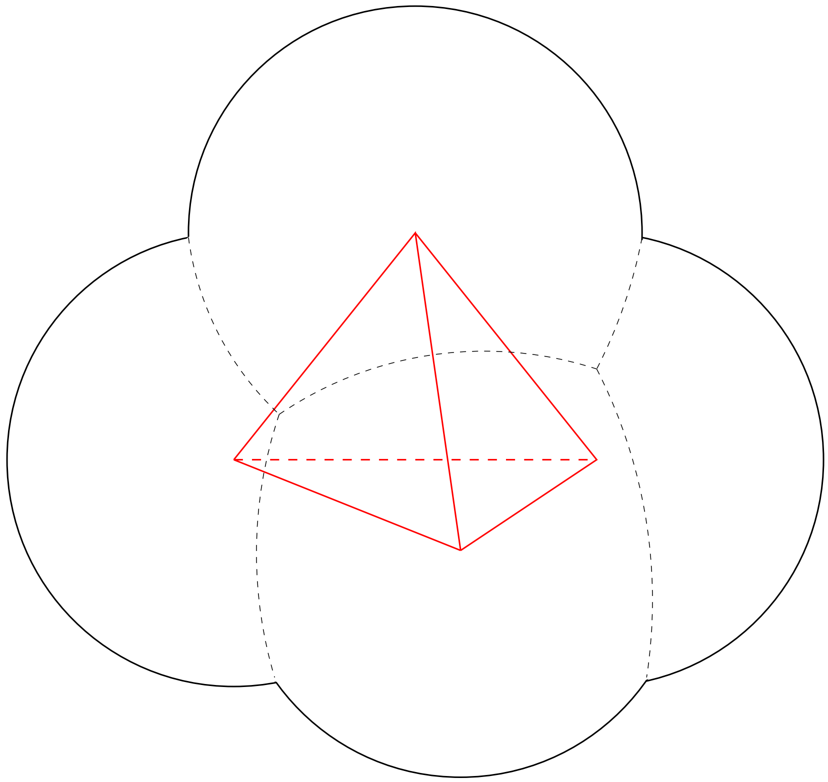


Figure 6.1: 4-Venn diagram of four spheres in 3 dimensions, showing the tetrahedron formed by the four curve centres in red

In the general case, Rényi, Rényi, and Surányi [100] proved that an independent family of $(n-1)$-spheres in n-dimensional space has at most $n+1$ members. Grünbaum's paper [56] contains some general geometric problems regarding independent families of convex higher-dimensional curves in higher-dimensional space, and mentions earlier work such as [22] and [100].

6.2 Diagrams in Higher Dimensions

In this section we appropriately generalize the definitions of Section 2.1 to (unbounded) m-dimensional Euclidean space, which we refer to as *m-space*. We follow Henle [73] in our definitions.

The coordinate system we use for m-space is the natural m-dimensional Cartesian coordinates with orthogonal axes $x_1, x_2, \ldots, x_m$, and any point is uniquely determined by specifying its location by these m coordinates. The definitions that follow overload the 2-dimensional terms where appropriate. As we noted in Chapter 2, the sphere in m-dimensional space is topologically equivalent to an $(m-1)$-dimensional plane and so is referred to as the $(m-1)$-sphere.

Definition. A *curve* is a subset of m-dimensional space that is deformable under continuous transformations to an $(m-1)$-sphere.

Other related terms from Section 2.1 such as *arc* and *vertex* can be easily generalized to their m-dimensional equivalents. As before, a *diagram* is a collection of curves in m-space.

A curve C partitions the Euclidean space into two disjoint sets, the bounded interior of C and the unbounded exterior, written $\mathit{interior}(C)$ and $\mathit{exterior}(C)$. The bounded interior of C is deformable under continuous transformations to a spherical region bounded by the $(m-1)$-sphere.

Definition. Let $\mathcal{D}$ be a diagram in m-space. A *region* of $\mathcal{D}$ is a maximal non-empty connected subset of the space containing no points on a curve.

Using topological terminology, a region in m-space is an *open m-cell*, and a curve is the boundary of an open m-cell.

Definition. Let $D = \{C_1, \ldots, C_n\}$ be a set of n curves in m-space. Then D is an *m-dimensional independent family* if each of the 2^n sets $X_1 \cap X_2 \cap \ldots \cap X_n$ is non-empty, where $X_i \in \{\mathit{interior}(C_i), \mathit{exterior}(C_i)\}$.

Definition. Let $D = \{C_1, \ldots, C_n\}$ be an m-dimensional independent family. If each of the 2^n sets $X_1 \cap X_2 \cap \ldots \cap X_n$ is connected, where $X_i \in \{\mathit{interior}(C_i), \mathit{exterior}(C_i)\}$, then D is an *m-dimensional Venn diagram.*

As in the planar case, there is a one-to-one correspondence between regions and the 2^n subsets $X_1 \cap X_2 \cap \ldots \cap X_n$ in an m-dimensional n-Venn diagram.

Definition. An m-dimensional diagram is *simple* if no point is shared by more than m curves.

For example, the diagram in Figure 6.1 is simple, as any three of the spheres share a common intersection point, but there is no point shared between all four of them.

A diagram in m-space can be embedded on the surface of the m-sphere (residing in $(m+1)$-space).

6.3 Venn Diagrams in Higher Dimensions

In this section we show the existence of simple Venn diagrams on spheres in higher dimensions, by a simple construction.

Given m, consider the collection of m curves embedded on the $(m-1)$-sphere $\{C_1, C_2, \ldots, C_m\}$, where

$$C_i = \{(x_1, x_2, \ldots, x_m) \,:\, x_i = 0 \text{ and } x_1^2 + x_2^2 + \ldots + x_{i-1}^2 + x_{i+1}^2 + \ldots + x_m^2 = 1\}.$$

We refer to this collection of m curves in $m-1$ dimensions as Δ_m.

For an example, Figure 6.2 shows the diagram Δ_3, which consists of three orthogonal great circles, each of which is a 1-sphere, embedded on the 2-sphere.

The following lemma is now easy to prove, since the m curves in Δ_m divide the surface of the $(m-1)$-sphere into 2^m regions.

Lemma 24. *Diagram Δ_m is a simple $(m-1)$-dimensional Venn diagram on the $(m-1)$-sphere.*

Proof. The proof showing that it is a m-Venn diagram consists of showing that there is a one-to-one correspondence between the 2^n subsets of the n-set and the regions of the diagram. Let $S_m = \{(x_1, \ldots, x_m) \,:\, x_1^2 + x_2^2 + \cdots + x_m^2 = 1\}$ be the $(m-1)$-sphere.

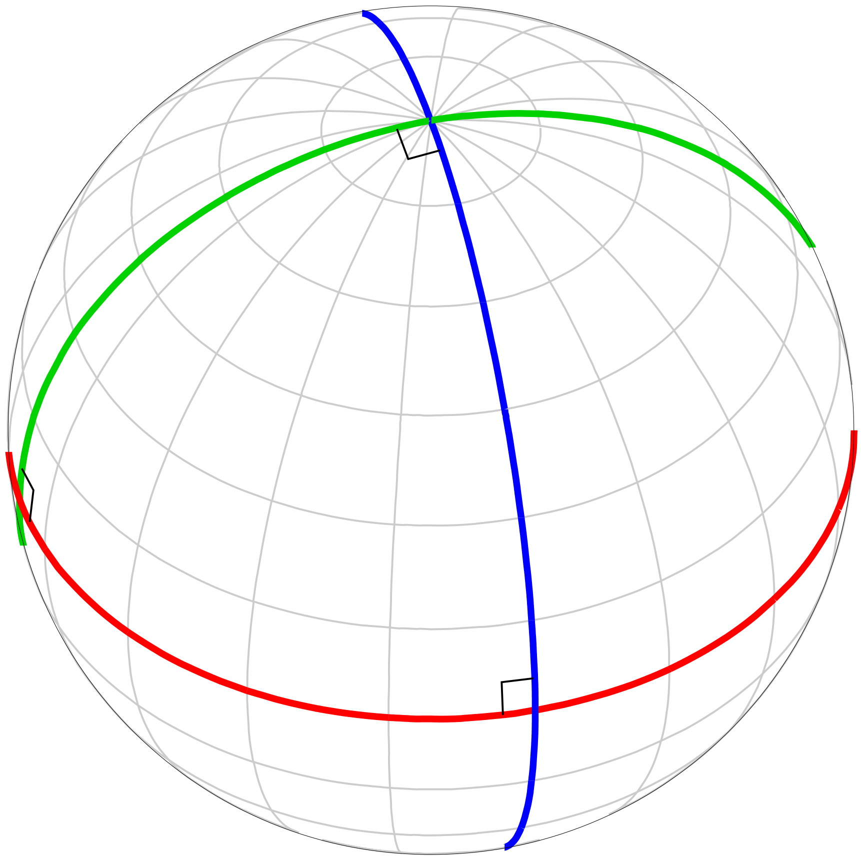


Figure 6.2: Diagram Δ_3, composed of three orthogonal 1-spheres

Consider an assignment of inside/outside to the two disjoint halves of the surface created by partitioning it by one of the m curves. Curve i divides the surface of the $(m-1)$-sphere into two halves, the half in which the coordinate x_i is positive and the half in which the coordinate x_i is negative. Without loss of generality, assume that the positive x_i half is the interior of curve C_i; *i.e* $\mathit{interior}(C_i) = \{(x_1, x_2, \ldots, x_m) \in S_m : x_i > 0\}$, and $\mathit{exterior}(C_i)$ is defined similarly.

Consider any subset v of the m-set $\{1, 2, \ldots, n\}$. To map this to a region of the diagram, consider the region $R \subset S_m = X_1 \cap X_2 \cap \cdots \cap X_m$ where $X_i = \mathit{interior}(C_i)$ if $i \in v$, $\mathit{exterior}(C_i)$ otherwise. The region R is connected since each $X_i \in \{\mathit{interior}(C_i)\ , \mathit{exterior}(C_i)\}$ is connected, and the intersection of connected subsets is connected. Also note that by the construction R is non-empty, as it is easy to find points $x = (x_1, x_2, \ldots, x_m)$ that satisfy $x \in S_m$ and $x_i > 0$ iff $X_i = interior(C_i)$, $x_i < 0$ otherwise. Any subset v of the m-set corresponds to such a region, and by the construction there are 2^m such regions, so Δ_m is an m-Venn diagram.

Finally, diagram Δ_m is simple. Consider that the $m-1$ curves $C_1, C_2, \ldots, C_{m-1}$ intersect only at the two points $(0, 0, 0, \ldots, 0, \pm 1)$, since those are the only points satisfying $x_1^2 + x_2^2 + \ldots + x_{i-1}^2 + x_{i+1}^2 + \ldots + x_m^2 = 1$ for $1 \leq i < m$; however these two points do not lie on curve C_m. A similar argument for the $m-1$ other possible sets of curves $\{C_1, C_2, \ldots, C_{i-1}, C_{i+1}, \ldots, C_m\}$ and their excluded curves C_i, with $1 \leq i < m$, shows that there is no point common to all of their intersections, and so the maximum number of curves sharing a single point is $m-1$. Since Δ_m is $(m-1)$-dimensional this shows that it is simple. □

Theorem 9. *There exists a simple $(m-1)$-dimensional n-Venn diagram on the $(m-1)$-sphere for $m \geq 1$ and $n \leq m$.*

Proof. The proof is constructive. For any $n = m$, diagram Δ_m is a simple $(m-1)$-dimensional n-Venn diagram on the $(m-1)$-sphere.

For $n-1 < m$, we remove a curve from Δ_m. If curve C_m is removed, the argument proving Lemma 24 applies in the same way if curve C_m is ignored, so the resulting diagram is a simple $(m-1)$-Venn diagram. Thus for any $n < m$ we can proceed inductively in the same way until only n curves remain; the resulting diagram will be a simple n-Venn diagram. □

Theorem 10. *There exists a simple m-dimensional n-Venn diagram for any $m \geq 2$ and $n \geq 1$.*

Proof. (A version of this proof is implied in [95])

For the case of $m = 2$ several simple constructions exist ([37, 38, 125], also some of the constructions discussed in Section 3.2.1). The proof proceeds inductively in the

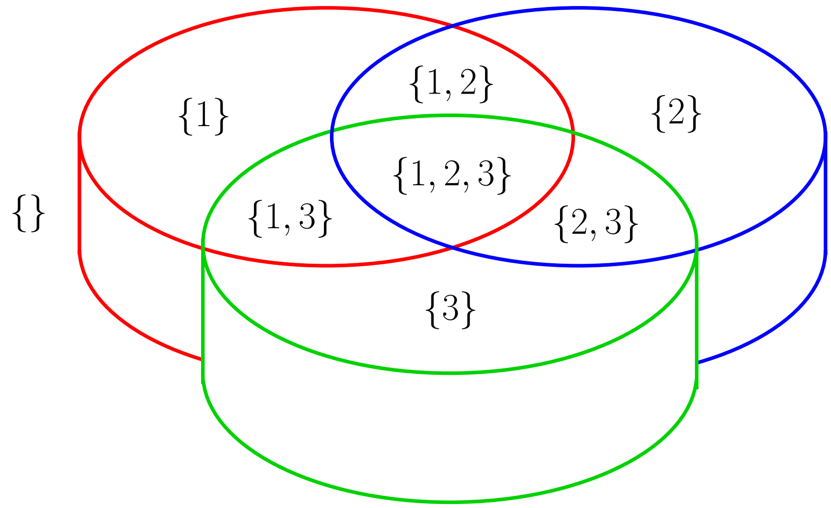


Figure 6.3: Extrusion of 3-circle Venn diagram of Figure 2.3 into three dimensions

dimension. Let $D = \{C_1, C_2, \ldots, C_n\}$ be a simple m-dimensional n-Venn diagram. Extrude all curves in D into the $(m+1)$st dimension to give $D' = \{C'_1, C'_2, \ldots, C'_n\}$ by treating each curve as the base of an $(m+1)$-dimensional prism, with all prisms having the same height (extrusion depth) ϵ, so

$$C'_i = \{(x_1, x_2, \ldots, x_m, x_{m+1} \in \mathbb{R}^{m+1} \ : \ (x_1, x_2, \ldots, x_m) \in C_i \text{ and } 0 \leq x_{m+1} \leq +\epsilon\} \ .$$

D' is then a simple $(m+1)$-dimensional n-Venn diagram. □

Example. As an example of the extrusion process in Theorem 10, see Figure 6.3, which shows the simple 3-circle Venn diagram of Figure 2.3 extruded into three dimensions to form three intersecting cylinders; this gives a 3-Venn diagram in three dimensions.

6.4 Realizing Groups in Higher Dimensions

There is a branch of mathematics called "representation theory" that studies group structures by representing their elements as linear transformations of vector spaces. Thus, care must be taken when we use the word "represent" when discussing examples of specific groups; generally we avoid the term in such contexts except when our use corresponds with similar usage in representation theory.

Recall the examples of direct product groups from Section 2.4.1. In this section we will be discussing a specific infinite family of direct product groups with important representations as bases of vector spaces.

This section was motivated by discussions with Brett Stevens.

Consider for any $k \geq 1$ the direct product group $\mathsf{C}_2 \times \mathsf{C}_2 \times \ldots \times \mathsf{C}_2 = \mathsf{C}_2^k$, which has order 2^k. Recall from Section 2.4.4 how direct product groups of cyclic groups of order n can be represented as base-n vectors, where composition of the group items is equivalent to vector addition modulo n. In this case, composition of group items is vector addition of binary k-vectors, where addition of two k-vectors is equivalent to computing the exclusive or ($\oplus$).

The group of all symmetries C_2^k thus corresponds to the vector space $\mathbb{Z}_2^k$, the collection of binary vectors of length k. For example, the identity symmetry is represented as the all-zero k-vector $(0, 0, \ldots, 0)$, called the *empty vector*. A symmetry group C_2 residing as a subgroup of C_2^k corresponds to a unit vector (all positions 0 except for a 1 in the position of the single C_2), for example $(1, 0, 0, \ldots, 0)$ together with the identity symmetry given by the empty k-vector, corresponds to the one-dimensional subspace given by that linear unit vector as a basis.

Compositions of two different unit vectors give a two-dimensional subspace, corresponding to four different symmetries (the identity, each symmetry corresponding to each unit vector alone, and the composition of the two). These two unit vectors then form a basis for the full two-dimensional subspace. In general, a collection of j distinct unit vectors will form a basis for the subspace of dimension 2^j that is $\mathbb{Z}_2^j$, and thus the set of k distinct unit vectors give a basis for $\mathbb{Z}_2^k$ itself.

We now investigate the types of symmetry operations possible on Δ_m, defined in Section 6.3. First, we observe that the types of isometries on Δ_m are severely restricted by its structure. An *axial* hyperplane is a hyperplane that contains all but one of the coordinate axes (and its normal vector is parallel to the axis it does not contain).

Lemma 25. *A total symmetry operation f on Δ_m must take the form $f(x_1, x_2, \ldots, x_m) = \pm x_1, \pm x_2, \ldots, \pm x_m$, that is, a composition of reflections across the axial m-dimensional hyperplanes.*

Proof. Consider any two curves $C_i, C_j \in \Delta_m$ where $i \neq j$. Consider two points $x \in C_i$ and $y \in C_j$ so that both x and y are contained on no other curves, *i.e* $x = (x_1, x_2, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_m)$ with $x_k \neq 0$ for $k \neq i$, and $y = (y_1, y_2, \ldots, y_{j-1}, 0, y_{j+1}, \ldots, y_m)$ with $y_k \neq 0$ for $k \neq j$. Any total symmetry operation on C_i that is not the identity and that preserves $y \in C_j$ while mapping x to a point $x' \neq x \in C_i$ is the reflection across the hyperplane $x_j = 0$, defined by the function $F(x) : x_j \rightarrow -x_j$, since that is the only dimension which C_j has no component in. This mapping takes x to $x' = (x_1, x_2, \ldots, -x_j, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_m)$. Note that this operation will also affect all other curves with components in dimension x_j, by reflecting them across the same hyperplane. The same argument applies to C_j with respect to C_i, so the only symmetry operations possible are reflection across $x_j = 0$, reflection across $x_i = 0$, and their composition (rotation around the common hyperplane defined by $x_i, x_j = 0$).

This argument holds for any pair of curves, so the total set of symmetries is restricted to symmetries of the form $F(x) : x_j \rightarrow -x_j$ for $1 \leq j \leq m$ and their compositions, which give only symmetries of the form $F(x_1, x_2, \ldots, x_m) = (\pm x_1, \pm x_2, \ldots, \pm x_m)$.

□

The immediate consequence of Lemma 25 is the following theorem, which specifies the symmetry group of Δ_m.

Theorem 11.

$$S_T(\Delta_m) \;\cong\; \mathsf{C}_2^m \;,$$

that is, the total symmetry group of Δ_m is isomorphic to the direct product group of m 2-cycles represented by the vector space $\mathbb{Z}_2^m$ with the group operation $\oplus$.

Proof. By Lemma 25, any valid total symmetry f of Δ_m corresponds to an m-vector $f = (f_1, f_2, \ldots, f_m)$ where $f_i \in \{+, -\}$. Thus, a symmetry is represented by the coordinates that are complemented when the symmetry acts on a point. There are thus 2^m possible symmetry operations, including the identity $(+, +, +, \ldots, +)$. Each operation consisting of a single coordinate negation gives the symmetry group C_2, a two-cycle. These m two-cycles map to m unit vectors giving a basis for the vector space $\mathbb{Z}_2^m$, and the composition of these m two-cycles gives the direct product group C_2^m, isomorphic to the group $(\mathbb{Z}_2^m, \oplus)$. $\square$

For example, Δ_3 in Figure 6.2 has total symmetries given by the reflections across the three planes $x = 0$, $y = 0$, and $z = 0$, which can be represented as the three 3-vectors $(-, +, +)$, $(+, -, +)$, and $(+, +, -)$. Composing these three symmetries with each other and the identity $(+, +, +)$ give eight possible symmetries, from the identity $(+, +, +)$ to the inversion symmetry $(-, -, -)$, and so we can say $S_T(\Delta_3) \cong \mathsf{C}_2^3$, corresponding to $\mathbb{Z}_2^3$.

The previous discussion has been laying the preliminaries for the following result, which provides a synthesis of the subgroups of C_2^m with diagrams realizing them as symmetry groups.

Theorem 12. *Let $\mathsf{G} \leq \mathsf{C}_2^m$ be a subgroup of the direct product group C_2^m. Then there exists a simple m-Venn diagram D'_m on the $(m-1)$-sphere such that $S_T(D'_m) \cong \mathsf{G}$.*

Proof. The proof is by construction. The technique is to modify the simple m-dimensional Venn diagram Δ_m by adding small local perturbations, called *widgets*, to a particular curve to explicitly allow only the symmetries present in G. We assume that the symmetries in G, by Lemma 11, are represented by m-vectors, as previously described, with the group operation being $\oplus$ of the vectors.

If $\mathsf{G} = \mathsf{C}_2^m$ then it is realizable by Δ_m, as in Theorem 11. For all other cases, henceforth we assume that G is a proper subgroup of C_2^m.

By Lemma 25, we can represent any total symmetry $f \in \mathsf{G}$ as an m-vector $f = (f_1, f_2, \ldots, f_m)$ where $f_i \in \{+, -\}$. We call a symmetry f a *unit symmetry* if $f_i = -$ for some single $i \in \{1, \ldots, m\}$ and $f_j = +$ for $j \neq i$; thus a unit symmetry gives the two-cycle given by reflection in a single coordinate. Thus we write $f \in \mathsf{G}$ if there is a one-to-one mapping from the m-vector specifying f to an m-vector in G.

We will realize $\mathsf{G} < \mathsf{C}_2^m$ by modifying Δ_m by adding $|\mathsf{G}|$ widgets to a single curve in Δ_m to restrict the allowable symmetry operations of that curve. To select which curve, let C_i be any curve such that there is not a unit symmetry $f \in \mathsf{G}$ such that $f_i = -$ and $f_j = +$ for $j \neq i$. Note that C_i must exist, since if all m unit symmetries are present in G, then since G is a group all compositions of these unit symmetries are present, which gives the entire group C_2^m, which is a contradiction since we assume G is a proper subgroup of C_2^m.

Recall that C_i is given by the points $\{(x_1, x_2, \ldots, x_m) : x_i = 0 \text{ and } x_1^2 + x_2^2 + \ldots + x_{i-1}^2 + x_{i+1}^2 + \ldots + x_m^2 = 1\}$; it has a component in every dimension except dimension i.

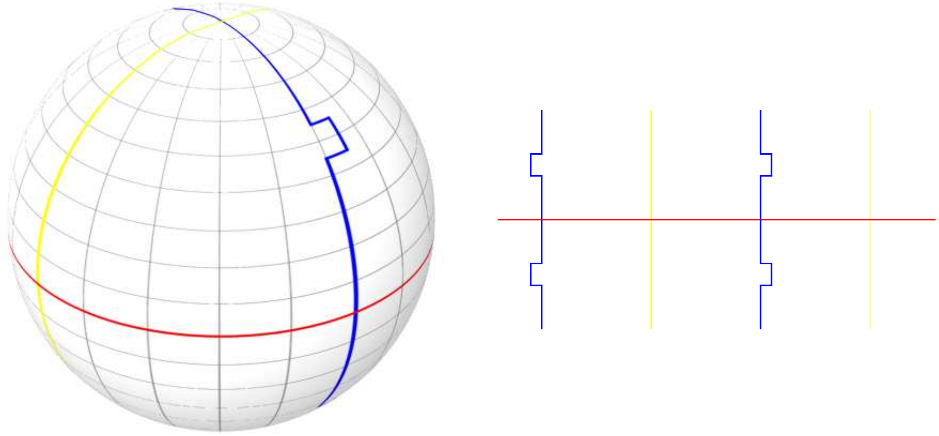
The widgets on curve C_i consist of an extrusion on the surface of the sphere of a section of C_i into dimension i, at the midpoints of the arcs between intersections with other curves; label these points $w(v)$ (for widget), where v will specify which point. The coordinates of point $w = (w_1, w_2, \ldots, w_m)$ are $w_j = \pm\frac{1}{\sqrt{m-1}}$ for $j \neq i$ and $w_i = 0$; there are thus 2^{m-1} such points w on C_i.

To specify a given w, given an m-vector v of $\pm$ signs, define the point $w(v)$ such that $w(v)_j = +\frac{1}{\sqrt{m-1}}$ if $v_j = +$ and $-\frac{1}{\sqrt{m-1}}$ otherwise. The coordinate v_i specifies the direction of the widget, as described next.

The design of the widget at point $w(v)$ is as follows. The curve C_i is an $(m-2)$-sphere residing on the $(m-1)$-sphere in m-space. Construct an $(m-3)$-sphere of fixed small radius ϵ centered at $w(v)$ on C_i. The coordinate v_i of v, since C_i does not have a component in the i direction, specifies the orientation of the widget as follows. Extrude the bounded surface contained inside this $(m-3)$-sphere a fixed small distance τ in the v_i direction in the i dimension ($i.e$ rotating it on the surface of the sphere in the positive i direction if $v_i = +$, the negative i direction otherwise) from C_i on the surface of the $(m-1)$-sphere, adding a connecting $(m-1)$-dimensional cylinder from this inner "plug" to the rest of C_i.

Example. As an example of the widget construction in three dimensions, see Figure 6.4. In this figure, a diagram derived from Δ_3 is shown, comprising three orthogonal curves on the 2-sphere with some widgets. We can assign axes so that the blue curve is C_3 and thus has the equation $x_1^2+x_2^2 = 1$. This curve has a widget at the centre of each edge between its crossing points with the yellow and red curves; the centre of the widget on the edge in x_1 and x_2's positive quadrant is at $w = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)$. The application of the construction specifies that the blue curve, which is a 1-sphere ($i.e$ circle), is cut in two places by a zero-sphere ($i.e$ two opposite points of distance ϵ from w along C_3) and the contained section is extruded in one direction (on the sphere).

For the construction, for each symmetry $f = (f_1, f_2, \ldots, f_m) \in \mathsf{G}$, with $f_i \in \{+, -\}$, add a widget to curve C_i at the point $w(f) = (w_1, w_2, \ldots, w_m)$ where, as specified earlier, $w_j = +\frac{1}{\sqrt{m-1}}$ if $f_j = +$ and $-\frac{1}{\sqrt{m-1}}$ otherwise, for $j \neq i$. The widget associated with f is extruded in the positive i direction if $f_i = +$ and in the negative i direction otherwise. Thus, each of the $|\mathsf{G}|$ symmetries gives exactly one widget. These widgets are all distinct (they are in distinct positions on C_i) since each symmetry has a unique set of the coordinates which are reflected under their



(a) Diagram on the sphere (b) Cylindrical projection of the diagram

Figure 6.4: 3-Venn diagram Δ_3 with example of widgets

operation, and this coordinate set is negative in the position vector $w(f)$. If $|\mathsf{G}| = 1$, *i.e* G consists solely of the identity symmetry given by the vector f with positive values for each dimension, then there is only one widget, at the position with all positive coordinates.

Applying this construction to Δ_m gives a modified diagram Δ'_m where all curves are identical to their counterparts in Δ_m except C_i, which has $|\mathsf{G}|$ widgets placed on it.

Example. Continuing the example of the widget construction in three dimensions in Figure 6.4, recall that the blue curve has been labelled C_3 for convenience. The symmetry group that is realized in this example contains the symmetries $\{(-,+,+), (+,-,+), (-,-,+), (+,+,+)\}$, which thus corresponds to the four widgets shown on C_3. As the axis labellings are arbitrary, we can say that the identity symmetry corresponds to the widget on the upper right-hand edge in the cylindrical projection. The unit symmetries $(-,+,+)$ and $(+,-,+)$ correspond to the two widgets reflected across the yellow and red curves from the identity symmetry widget, and the composition of these two symmetries gives the bottom left-hand widget. In this example the direction of the initial widget's extrusion from C_3 is in the positive x_3 direction, since the symmetries in the group represented all have $x_3 = +$.

We now prove that $S_T(\Delta'_m) \cong \mathsf{G}$; that is, that the total symmetry group of Δ'_m realizes the group G. The proof continues in two stages; first, to show that any symmetry in G is a total symmetry for Δ'_m, and second, to show that applying any symmetry not in G is not a symmetry for Δ'_m.

For the first part, any curve C_j with $j \neq i$ is preserved under any total symmetry in C_2^m, by Theorem 11. Similarly, any point on C_i that is not part of a widget maps onto another point on C_i that is not on a widget. By the symmetry of the widgets and the symmetry of their placement on C_i, any point not within ϵ of a widget cannot be mapped onto a widget, and vice versa. It remains to argue under any symmetry $f \in \mathsf{G}$ any widget on C_i is mapped to another widget, thus preserving the symmetry.

First note that any widget at location $w(v)$ for some vector v that is constructed in the positive i direction (respectively, negative i direction) will never be mapped by a symmetry $f \in \mathsf{G}$ onto a widget at the same location but in the negative i direction (respectively, positive i direction) since G does not contain the unit vector mapping to a symmetry that reflects widgets only in the i dimension; this is how C_i was chosen in the construction.

Every group G contains the identity symmetry, so there is always one widget that is constructed at $w(\ (+, +, \ldots, +)\ )$, *i.e* the location with all coordinates positive. Any other widget $w(f)$ at some location on C_i corresponds exactly to some total symmetry operation $f \in \mathsf{G}$ by the construction, and so applying the symmetry f to $w(\ (+, +, \ldots, +)\ )$ maps it onto the widget $w(f)$ (and vice versa, since every symmetry composed with itself gives the identity). Applying another symmetry $g \neq f \in \mathsf{G}$ to the widget at $w(f)$ gives the position $w(h)$ corresponding to the symmetry composed of the composition $h = g \cdot f$, and since $h \in \mathsf{G}$ as G is closed under composition there must be a widget at $w(h)$ by the construction. Since any symmetry can be reached by composition, any symmetry maps any widget onto some other widget, which must already exist by the construction. So since all widgets map onto other widgets and all points that are not on widgets map onto other points that are not on widgets, all curves are preserved under any symmetry.

For the second part, we argue that applying any symmetry not corresponding to a vector in G maps a widget on C_i to a location on C_i where there is no widget, and thus that symmetry is not a symmetry for Δ'_m. Let $d \in \mathsf{C}_2^m$ be a vector such that $d \notin \mathsf{G}$. Consider the widget $w(\ (+,+,\ldots,+)\ )$ corresponding to the identity symmetry, located at the widget position with all positive coordinates, extruded in the positive i direction. Applying d to this widget maps it to a point on C_i where no widget has been placed, since if $d(w(\ (+,+,\ldots,+)\ ))$ had a widget in that position, it could only be because of a symmetry in G that the construction had put it there. Thus d is not a valid symmetry for C_i and thus not for Δ'_m.

Thus, since C_i only maps onto itself under exactly the symmetries in G, and all other curves map onto themselves under any symmetry in C_2^m, the full diagram Δ'_m has exactly the total symmetry group allowable by the modified C_i, which is G. □

6.5 Curve-Preserving Symmetry and Open Questions

A natural extension of the previous discussion is to consider the curve-preserving symmetry groups possible in higher dimensions. In three dimensions, Δ_3 in Figure 6.2 has the full oriented curve-preserving symmetry group isomorphic to the symmetric group P_3, all permutations of the the 3-set.

Lemma 26.

$$S_{OC}(\Delta_3) \cong \mathsf{P}_3.$$

Proof. This is easy to see, since, first, there are reflections across planes that are orthogonal to one curve and bisect the angle between the planes containing the other two curves that swap those two curves and preserve their orientations. There are also rotations about about an axis through a face that rotate all three curves while preserving orientation, and combining these two operations can permute the three curves in all possible ways while preserving orientation. □

It seems natural to consider this in any number of dimensions, giving us:

Lemma 27.

$$S_{OC}(\Delta_m) \cong \mathsf{P}_m, \text{ or has a subgroup isomorphic to it.}$$

Proof. The idea behind the proof of this is that any point on the $(m-1)$-sphere is given by coordinates in $(x_1, x_2, \ldots, x_m)$, and by swapping coordinates we can interchange any set of curves while preserving interiors and exteriors, which give the orientation. The group is isomorphic to at least P_m since there may be other curve-preserving symmetries that occur in higher dimensions. □

It seems plausible that a widget-based construction similar to the one used in the proof of Theorem 12 could be used to generate a simple m-Venn diagram on the $(m-1)$-sphere realizing any particular subgroup of P_m. Thus we offer the following conjecture:

Conjecture 3. *Let $\mathsf{G} < \mathsf{P}_m$ be a subgroup of the symmetric group P_m. Then there exists a simple m-Venn diagram D_m on the $(m-1)$-sphere such that $S_C(D_m) \cong \mathsf{G}$.*

If we now consider ignoring orientation of curves, then we find that the unoriented curve-preserving symmetry group of Δ_3 is the same as that of the regular cube in three dimensions, as we will discuss later in Section 8.3.2.

Observation 4. $S_C(\Delta_3) = \mathsf{O}_h \cong \mathsf{P}_4 \times \mathsf{C}_2$

Proof. Consider embedding the diagram on the surface of a cube, with edges of the diagram orthogonal to the edges of the cube, as in Figure 6.5, and vertices of the cube in the centre of each face. Clearly any symmetry of the cube, including the reflective symmetries, is also a curve-preserving symmetry for the diagram, and the unoriented curve-preserving symmetry group of the cube is O_h. $\square$

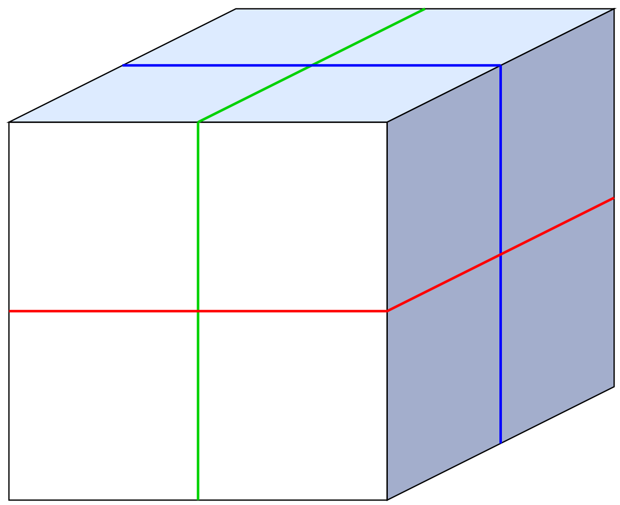


Figure 6.5: Simple 3-Venn diagram on the cube

Compare Figure 6.5 to Figure 8.14 in Section 8.3.2, which shows the same diagram on the corresponding dual solid, the octahedron.

Again it seems natural to conjecture the following:

Conjecture 4. *For all $m \geq 2$,*

$$S_C(\Delta_m) \cong \mathsf{P}_m \times \mathsf{C}_2,$$

or has a subgroup isomorphic to it.

It may be tempting to try to generalize Theorem 12 to oriented curve-preserving symmetries, to show that any subgroup of the $\mathsf{P}_m \times \mathsf{C}_2$ is realizable by a simple m-Venn diagram on the $(m-1)$-sphere in m dimensions. However, consider a subgroup consisting of some oriented curve-preserving symmetries and all of the total symmetries, which must include all oriented total symmetries that leave all curves unchanged except one; namely a reflection across the hyperplane containing that curve—for example, in Figure 6.2 the three reflections across the three orthogonal planes containing each of the three curves must be included. However, it is impossible to then use a widget-based construction to exclude some curve-preserving symmetries, as adding any widget to a curve will destroy the total symmetries just described. For example, in Figure 6.4, the widgets on the blue curve will destroy the total symmetry given by the reflection across the plane containing the blue curve. This is because a widget must deviate in some way from that plane the curve is contained in, and this deviation will reflect across the plane under that reflection to some place without a widget. These considerations seem to preclude any widget-based construction for subgroups of the full group of possible oriented curve-preserving symmetries.

Finally, an interesting question is whether the example of Figure 6.1 be generalized to m dimensions, so that $m+1$ $(m-1)$-balls can be arranged with their centres at the corners of a generalized tetrahedron to form an $(m+1)$-curve Venn diagram in m-space.

Chapter 7

Shift Register Sequences and Rotary Reflection Symmetries

This chapter discusses an interesting connection between certain equivalence classes of strings under rotation, and a method of constructing diagrams on the sphere that have a natural type of rotary reflection symmetry. We introduce a new framework for these new types of n-Venn diagram, which turn out to have unoriented curve-preserving symmetry group S_{2n} when n is a power of two, prime, or a Poulet number (defined below). As in the construction of [54] from Section 3.3.2, the diagrams can be built from a fundamental domain of the dual, built by finding representatives of classes that decompose the boolean lattice. As in previous sections we provide a few diagrams as motivation, from which we extract the fundamental domain and generalize the appropriate properties to find more diagrams. First, however, we discuss some properties of different types of shift register sequences that will prove useful when constructing the diagrams.

7.1 Necklaces and Cycling Shift Register Sequences

In this section we define and discuss some basic properties of some important ways of decomposing the set of all n-bit binary strings B_n. The first two functions we examine have been defined and discussed in [51], and we also introduce the inverting complementing cycling register. A series of lemmas lead up to an equivalence result that will prove useful in a construction in following sections.

We first review the *pure cycling register*, defined as a function on n-bit strings, from Section 3.1.2.

Definition ([51, p. 171])**.** Given a bitstring $\beta = \beta_1\beta_2\cdots\beta_n$, the *pure cycling register* (PCR) is a bijection $PCR_n : B_n \to B_n$ defined as $PCR_n(\beta) = \beta_2\beta_3\cdots\beta_n\beta_1$, or simply $PCR(\beta)$.

Recall from our earlier use that $PCR^i(\beta) = PCR(PCR^{i-1}(\beta))$, where $PCR^1(\beta) = PCR(\beta)$. We also have the relation $\overset{PCR}{\sim}$ on B_n by $x \overset{PCR}{\sim} y$ iff $y = PCR^i(x)$ for some $i \geq 1$. Then $\overset{PCR}{\sim}$ is an equivalence relation on B_n, called the *necklace* relation, and the equivalence classes are called *necklaces*.

The necklace poset is well-studied from the perspective of shift register sequences. Several algorithms exist for generating representatives from each necklace class; see [44, 45, 105], and [46, 51] for background from the perspective of shift register sequences.

Given n, the number of necklaces is

$$N(n) \ = \ \frac{1}{n}\sum_{d|n}\phi(d)2^{n/d} \ , \tag{7.1}$$

where ϕ is the Euler ϕ-function (or *totient* function), defined as

$$\phi(x) = |\ \{y : y \leq x \text{ is relatively prime to } x\}\ |.$$

We used the PCR and necklaces in Section 3.1.2 to understand the labellings of the fundamental domains of dual graphs of planar rotationally symmetric diagrams.

Definition. Given a bitstring $\beta = \beta_1\beta_2\cdots\beta_n$, the *cycling complement register* (CCR) is a bijection $CCR_n : B_n \to B_n$ defined as $CCR_n(\beta) = \beta_2\beta_3\cdots\beta_n\overline{\beta_1}$, or just $CCR(\beta)$.

n	PCR	CCR
	$N(n)$	$N^*(n)$
1	2	1
2	3	1
3	4	2
4	6	2
5	8	4
6	14	6
7	20	10
8	36	16
9	60	30
10	108	52

Table 7.1: Number of equivalence classes for CCR and PCR for small n

The CCR function is defined and described in [51]; see also [82, Sec. 4] for their relation to vortex-free tournaments, and [89, Appendix 1] for more detail on the number of classes the CCR decomposes B_n into.

The CCR function, though bearing obvious similarities to the traditional PCR function, has markedly different behaviour, which we will explore.

We can define the relation $\overset{CCR}{\sim}$ on B_n by $x \overset{CCR}{\sim} y$ iff $x = CCR_n^i(y)$ for some $i \geq 1$. Then $\overset{CCR}{\sim}$ is an equivalence relation on B_n which we call the *CCR necklace* relation, and the equivalence classes are called *CCR necklaces*. Given $x \in B_n$, the class $\{y : y \overset{CCR}{\sim} x, y \in B_n\}$ is written $[x]$, and the size of the class is written $m_x = |[x]|$.

As shown in [51], the number of equivalence classes that the CCR necklace function decomposes B_n into is

$$N^*(n) \;=\; \frac{1}{2}N(n) - \frac{1}{2n}\sum_{2d|n}\phi(2d)2^{n/2d}\;. \tag{7.2}$$

The numbers of classes for both of the functions is shown for small n in Table 7.1, taken from [51].

The following lemma was tersely proven in [15]; here we provide a more formal proof using the PCR and CCR notation.

Lemma 28 ([15]). *There is a bijection between CCR necklaces and odd weight necklaces.*

Proof. Consider the bitstrings $x = x_1\ x_2 \cdots x_n$ and $CCR(x) = x_2\ \ \cdots x_n\ \overline{x_1}$ and let $y = y_1 y_2 \cdots y_n = x \oplus CCR(x)$ such that $y_i = x_i \oplus x_{i+1}$ for $i < n$ and $y_n = x_n \oplus \overline{x_1}$.

The parity of the string y is found from the $\oplus$ of all of its bits; thus the parity of y is

$$\begin{aligned} y_1 \oplus y_2 \oplus \cdots \oplus y_n &= (x_1 \oplus x_2) \oplus (x_2 \oplus x_3) \oplus \cdots \oplus (x_n \oplus \overline{x_1}) \\ &= x_1 \oplus x_2 \oplus x_2 \oplus x_3 \oplus x_3 \oplus \cdots \oplus x_n \oplus x_n \oplus \overline{x_1}\ \text{, since } \oplus \text{ is associative} \\ &= x_1 \oplus \overline{x_1}\ \text{, since } a \oplus a = 0 \text{ and } b \oplus 0 = b \\ &= 1\ , \end{aligned}$$

and so y has odd weight.

Now $CCR(x) \oplus CCR^2(x) = PCR(y)$, since

$$\begin{aligned} CCR(x) \oplus CCR^2(x) &= x_2\ \ \cdots x_n\ \overline{x_1} \oplus x_3\ \ \cdots x_n\ \overline{x_1}\ \overline{x_2} \\ &= x_2 \oplus x_3, x_3 \oplus x_4, \cdots, x_n \oplus \overline{x_1}, \overline{x_1} \oplus \overline{x_2} \\ &= x_2 \oplus x_3, x_3 \oplus x_4, \cdots, x_n \oplus \overline{x_1}, x_1 \oplus x_2 \quad \text{since } a \oplus b = \overline{a} \oplus \overline{b} \\ &= PCR(y), \end{aligned}$$

by definition of the PCR. By induction, any $z \in B_n, z = CCR^i(x)$ for some $i \geq 0$ has $z = PCR^i(y)$. Without loss of generality, since x was arbitrarily chosen, let r be such that it is lexicographically least of all z such that $z \overset{PCR}{\sim} y$; then as per our earlier definition of necklace, r is the odd weight necklace corresponding to x. $\square$

This immediately gives the following corollary:

Corollary 7.

$$N^*(n) \; = \; |\{x \in N(n),\, |x| \; odd\}|.$$

This can also be proved by working directly from the proof of Equation 7.1, for example from [51].

Observe that for any $x \in B_n$, we have that $CCR^n(x) = \overline{x}$, and thus $CCR^{2n}(x) = x$. Thus the CCR necklaces have size at most $2n$, and the next lemma tells us that all CCR necklaces have this size exactly when n is a power of two. This lemma will have important implications for constructing Venn diagrams from posets of necklaces using the CCR.

Lemma 29. *Given n, then $n = 2^k$ for some $k \geq 0$ if and only if for all $x \in B_n$, $m_x = 2n$.*

Proof. If $n = 2^k$ for some $k \geq 0$, then applying Equation 7.2 and substituting for $N(n)$ we have

$$\begin{aligned} N^*(n) &= \frac{1}{2n} \sum_{d=2^i, i \geq 0} \phi(d) 2^{n/d} - \frac{1}{2n} \sum_{j=2^i, i>0} \phi(j) 2^{n/j} \\ &= \frac{1}{2n} 2^n \ . \end{aligned}$$

Since the maximum size of any one CCR necklace is $2n$ then the fact that there are only $2^n/2n$ of them means they must all have size $2n$ to include all 2^n bitstrings. For the proof's inverse direction we show an example of a class with size less than $2n$ by assuming $n \neq 2^k$. If n has some odd prime divisor k, consider the string $x \in B_n$ where $x = (0^{n/k} 1^{n/k})^{(k-1)/2}\ 0^{n/k}$; *i.e* $x = 0^{n/k}\ 1^{n/k}\ 0^{n/k}\ 1^{n/k}\ \ldots\ 0^{n/k}$. For example, if $n = 6$ then $k = 3$ and x is the string 001100. Now observe that $CCR^{2n/k}(x) = x$, and thus the class $[x]$ has size $m_x = 2n/k < 2n$. $\square$

Definition. Given a bitstring $\beta = \beta_1\beta_2\cdots\beta_n$, the *inverting cycling complement register* (*ICCR*) is a bijection $ICCR_n : B_n \to B_n$ defined as $ICCR_n(\beta) = \overline{\beta_2}\,\overline{\beta_3}\,\cdots\,\overline{\beta_n}\,\beta_1$, or just $ICCR(\beta)$.

For an example of the CCR versus ICCR classes, CCR_4 gives:

$$0000 \to 0001 \to 0011 \to 0111 \to 1111 \to 1110 \to 1100 \to 1000 \to$$

and

$$0010 \to 0101 \to 1011 \to 0110 \to 1101 \to 1010 \to 0100 \to 1001 \to ,$$

whereas $ICCR_4$ gives:

$$0000 \to 1110 \to 0011 \to 1000 \to 1111 \to 0001 \to 1100 \to 0111 \to$$

and

$$0010 \to 1010 \to 1011 \to 1001 \to 1101 \to 0101 \to 0100 \to 0110 \to .$$

7.2 Relationships between CCR and ICCR Classes

Careful readers will notice that the classes given in the examples in the previous section for CCR_4 and $ICCR_4$ contain the same strings as the two classes from the CCR, just in a different order. In this section, through a series of observations and lemmas, we establish necessary and sufficient conditions for the general version of this correspondence between the CCR necklaces and ICCR necklaces.

Observation 5.

$$ICCR(x) = \overline{CCR(x)}$$

and

$$CCR(x) = \overline{ICCR(x)}.$$

Lemma 30. *Let $x \in B_n$, and let k be such that the shift $CCR_n^k(x) = \overline{x}$. If n is even, then k is even.*

Proof. Since $CCR^n(x) = \overline{x}$ we may assume that $0 < k \leq n$. Assume n is even and k is odd and derive a contradiction. There are several cases:

Case 1: $\mathbf{k = n}$: Contradiction, since n is even.

Case 2: $\mathbf{k < n/2}$: In this case, write x as

$$x = x_1 x_2 \cdots x_k x_{k+1} \cdots x_{n-k+1} \cdots x_{n-1} x_n$$

and

$$CCR^k(x) = x_{k+1} \cdots x_{n-k+1} \cdots x_{n-1} x_n \overline{x_1}\ \overline{x_2} \cdots \overline{x_k}\ .$$

Let the substring $x_1 x_2 \cdots x_k = \kappa$, with $|\kappa| = k$. Then we can write the above as

$$\begin{array}{ccccccl} x & = & \kappa & x_{k+1} \cdots x_{n-k} & x_{n-k+1} \cdots x_{n-1} x_n & , & \text{and} \\ CCR^k(x) & = & x_{k+1} \cdots x_{2k} & x_{2k+1} \cdots x_{n-1} x_n & \overline{\kappa} & . & \end{array}$$

Since $CCR^k(x) = \overline{x}$, by the above $\overline{\kappa} = x_{k+1}\ \cdots\ x_{k+k}$, and $x_{n-k+1}\ \cdots\ x_n = \kappa$, and so $x = \kappa\ \overline{\kappa}\ x_{2k}\ \cdots\ x_{n-k}\ \kappa$. Substituting into the above, we have

$$\begin{array}{cccccccl} x & = & \kappa & \overline{\kappa} & x_{2k+1} \cdots x_{n-2k} & x_{n-2k+1} \cdots x_{n-k} & \kappa & , \quad \text{and} \\ CCR^k(x) & = & \overline{\kappa} & x_{2k+1} \cdots x_{3k} & x_{3k+1} \cdots x_{n-2k} & \kappa & \overline{\kappa} & . \end{array}$$

Continuing inductively on the bit position from either end of x, we finally have $x = \kappa\ \overline{\kappa}\ \kappa\ \cdots\ \overline{\kappa}\ \kappa$, or more compactly $x = (\kappa\ \overline{\kappa})^i\ \kappa$, for some $i \geq 0$. But then, since $|\kappa|$ is odd, $|x|$ is odd, a contradiction.

Case 3: $\mathbf{k = n/2}$: In this case, write x as

$$x = x_1 x_2 \cdots x_k x_{k+1} \cdots x_{n-k+1} \cdots x_{n-1} x_n$$

and

$$CCR^k(x) = x_{k+1} \cdots x_{n-k+1} \cdots x_{n-1} x_n \overline{x_1}\ \overline{x_2} \cdots \overline{x_k}\ .$$

Let the substring $x_1 x_2 \cdots x_k = \kappa$, with $|\kappa| = k$. Then we can write the above as

$$\begin{array}{rcccl} x & = & \kappa & x_{k+1} \cdots x_n & , \quad \text{and} \\ CCR^k(x) & = & x_{k+1} \cdots x_n & \overline{\kappa} & . \end{array}$$

Since $CCR^k(x) = \overline{x}$, by the second line of the above, $\overline{\kappa} = x_{k+1}\ \cdots\ x_n$, but the second half of the first line gives $\overline{\overline{\kappa}} = \kappa = x_{k+1}\ \cdots\ x_n$, which is a contradiction.

Case 4: $\mathbf{k > n/2}$: The argument is similar to the case for $k < n/2$. Write x as

$$\begin{array}{rcllll} x & = & x_1\ x_2 \cdots & \cdots x_k & x_{k+1} \cdots x_{n-1}\ x_n & , \quad \text{and} \\ CCR^k(x) & = & x_{k+1}\ x_{k+2} \cdots x_n & \overline{x_1}\ \overline{x_2} \cdots \overline{x_{2k-n}} & \overline{x_{2k-n+1}} \cdots \overline{x_k} & . \end{array}$$

Let the substring $x_1 x_2 \cdots x_{2k-n} = \kappa$, with $|\kappa| = k - (n-k) = 2k - n$. Since $CCR^k(x) = \overline{x}$, we have

$$\begin{array}{rclllll} x & = & & \kappa & \cdots x_k & x_{k+1} \cdots x_{n-1}\ x_n & , \quad \text{and} \\ CCR^k(x) & = & x_{k+1}\ x_{k+2} \cdots x_n & \overline{\kappa} & & \overline{x_{2k-n+1}} \cdots \overline{x_k} & , \end{array}$$

and thus

$$\begin{array}{rcllll} x & = & \kappa & \cdots x_k & \overline{\kappa} \cdots x_{n-1}\ x_n & , \quad \text{and} \\ CCR^k(x) & = & \overline{\kappa} \cdots x_n\ \ \overline{\kappa} & & \overline{x_{2k-n+1}} \cdots \overline{x_k} & , \end{array}$$

and thus

$$\begin{aligned} x &= \kappa \;\; \overline{\kappa} \qquad\quad \cdots x_k \;\; \overline{\kappa} \;\; \cdots x_{n-1}\; x_n \;\; , \qquad \text{and} \\ CCR^k(x) &= \overline{\kappa} \;\; \cdots x_n \;\; \overline{\kappa} \qquad\quad \overline{\kappa} \;\; \cdots \overline{x_k}\; . \end{aligned}$$

Continuing in this fashion on the bit position we finally have $x = \kappa\; \overline{\kappa}\; \kappa\; \cdots\; \kappa\; \overline{\kappa}\; \overline{\kappa}\; \kappa\; \overline{\kappa}\; \cdots\; \overline{\kappa}$, or more compactly $x = (\kappa\; \overline{\kappa})^i\; (\overline{\kappa}\; \kappa)^j \overline{\kappa}$, with $|(\kappa\; \overline{\kappa})^i| = k$, but this is a contradiction as k was assumed to be odd.

□

Corollary 8. *Let $x \in B_n$. If n is even, then m_x is even.*

Proof. Let $x \in B_n$, and let k' be such that the shift $CCR_n^{k'}(\overline{x}) = x$. We can apply the same argument as in the proof of Lemma 30 to prove that if n is even, k' must also be even. Then the size of the CCR equivalence class for x is $k + k'$, as $CCR_n^{k+k'}(x) = x$, which is even. □

Similarly, let k be such that, for any $x \in B_n$, the shift $CCR_n^k(x) = \overline{x}$. Then $k = m_x/2$.

Lemma 31. *Let $x \in B_n$, and let k be such that the shift $ICCR_n^k(x) = \overline{x}$. If n is even, then k is even.*

Proof. The proof is very similar to the proof of Lemma 30, using Observation 5 where appropriate. □

Lemma 32. *Let $x \in B_n$. Let k be such that the shift $CCR_n^k(x) = \overline{x}$, and let k' be such that the shift $ICCR_n^{k'}(x) = \overline{x}$. Then if n is even, $k = k'$.*

Proof. Lemmas 30 and 31 tell us that k and k' are both even. Assume $k < k'$. Then

$\overline{x} = CCR^k(x)$. Consider that

$$\begin{aligned} ICCR(x) &= \overline{x_2} \cdots \overline{x_n}\ x_1, \quad \text{and} \\ ICCR^2(x) &= x_3 \cdots x_n\ \overline{x_1}\ \overline{x_2} \\ &= CCR_n^2(x)\ . \end{aligned} \tag{7.3}$$

and by repeated application $CCR^i(x) = ICCR^i(x)$ with i even. Then $\overline{x} = CCR^k(x) = ICCR^k(x)$, contradicting that $k < k'$. A similar contradiction occurs if $k' < k$. □

Corollary 9. *For $x \in B_n$, if n is even, then the size of the CCR and ICCR equivalence classes are equal.*

Proof. Similar to the proof of Corollary 8. □

Thus, if n is even, we can denote the size of the CCR or ICCR necklaces as m_x.

Lemma 33.

$$CCR^{m_x/2+1}(x) = ICCR(x).$$

Proof. Let $x = x_1\ x_2\ \cdots x_n$. Then

$$\begin{aligned} CCR(x) &= x_2 \cdots x_n\ \overline{x_1}, \\ CCR^2(x) &= x_3 \cdots x_n\ \overline{x_1}\ \overline{x_2}, \\ &\vdots \\ CCR^{m_x/2}(x) &= \overline{x_1} \cdots \overline{x_{n-1}}\ \overline{x_n} = \overline{x}, \quad \text{and} \\ CCR^{m_x/2+1}(x) &= \overline{x_2} \cdots \overline{x_n}\ x_1, \\ &= ICCR(x). \end{aligned}$$

□

Corollary 10.

$$CCR^{m_x/2+1}(x) \;=\; \overline{CCR(x)}.$$

If n is even, we can now prove, using all of the previous lemmas, that there is a complete bijection from the CCR function equivalence class of any $x \in B_n$ to the ICCR function equivalence class. This lemma will have important consequences when we are constructing Venn diagrams using the CCR necklaces.

Lemma 34. *If n is even, then for any $x \in B_n$, $\{y : y \overset{\text{CCR}}{\sim} x\} = \{z : z \overset{\text{ICCR}}{\sim} x\}$,* i.e *the equivalence classes of B_n under the ICCR function are the same classes as under the CCR function.*

Proof. Given a string $x = x_1\, x_2 \cdots x_n$, first begin with Equation 7.3 and by repeated application we have that $CCR^i(x) = ICCR^i(x)$ for $i < m_x$, with i even, up to $CCR^{m_x}(x) = ICCR^{m_x}(x) = x$, as in the proof of Lemma 32. Note that, from Corollary 9, the CCR and ICCR equivalence classes are the same size, m_x. This covers the even shifts of x.

Using induction and applying Equation 7.3 and Lemma 33 repeatedly, we have

$$\begin{aligned}
ICCR(x) &= CCR^{m_x/2+1}(x) \\
ICCR^3(x) &= CCR^{m_x/2+3}(x) \\
&\vdots \\
ICCR^{m_x/2+1}(x) &= CCR^{m_x+1}(x) = CCR(x) \\
ICCR^{m_x/2+3}(x) &= CCR^3(x) \\
&\vdots \\
ICCR^{m_x-1}(x) &= CCR^{m_x/2-1}(x) \text{ , and finally returning to} \\
ICCR(x) &= CCR^{m_x/2+1}(x) \text{ , as in Lemma 33,}
\end{aligned}$$

where addition is modulo m_x. This covers the odd shifts of x.

This exhausts the equivalence classes concerned, and so we have, given x,

$$CCR_n^i(x) = \begin{cases} ICCR^i(x) & \text{if } i \text{ even} \\ ICCR^{(i+m_x/2) \pmod{m_x}}(x) & \text{if } i \text{ odd} \end{cases}$$

for $0 \leq i < m_x$, and

$$ICCR_n^i(x) = \begin{cases} CCR^i(x) & \text{if } i \text{ even} \\ CCR^{(i-(m_x/2)) \pmod{m_x}}(x) & \text{if } i \text{ odd} \end{cases}$$

for $0 \leq i < m_x$, where in both cases recall that $CCR^0(x) = ICCR^0(x) = x$. □

Lemma 35. *The equivalence classes of B_n under the ICCR are the same as under the CCR if and only if n is even.*

Proof. The sufficient condition was shown in Lemma 34.

For necessity, prove by contradiction by assuming n is odd and showing that there is a string that belongs to distinct classes under the *CCR* and the *ICCR*. Consider the string

$$x = 010101 \cdots 10$$

and consider the *CCR* necklace it belongs to.

$$\begin{aligned} CCR(x) &= 10101 \cdots 01 \text{ , and} \\ CCR^2(x) &= 0101 \cdots 10 = x \text{ ,} \end{aligned}$$

and thus $|\{y \in B_n, y \overset{CCR}{\sim} x\}| = 2$. But under the *ICCR* we have that

$$ICCR(x) = 0101 \cdots 010 = x,$$

and thus $|\{y \in B_n, y \overset{CCR}{\sim} x\}| = 1$, and so these two classes are distinct, giving a contradiction. ☐

Corollary 11. *Let $x \in B_n$. If $n = 2^k$ for some k, then $m_x = |[x]| = 2n$ under $\overset{\text{ICCR}}{\sim}$.*

Proof. By Lemma 35 and Lemma 29. ☐

Observation 6. *Let $x \in B_n$. If n is even, then $CCR^n(x) = ICCR^n(x) = \overline{x}$.*

Lemma 36. *Let $x \in B_n$. If n is odd, then $|m_x| \leq n$, where $m_x = |[x]|$ under $\overset{\text{ICCR}}{\sim}$.*

Proof. Consider applying the ICCR for any even number of shifts, as

$$\begin{aligned}
ICCR(x) &= \overline{x_2}\,\overline{x_3}\,\ldots\,\overline{x_n}\,x_1 \\
ICCR^2(x) &= x_3\,x_4\,\ldots x_n\,\overline{x_1}\,\overline{x_2} \\
ICCR^3(x) &= \overline{x_4}\,\overline{x_5}\,\ldots\,\overline{x_n}\,x_1\,x_2\,x_3 \\
ICCR^4(x) &= x_5\,\ldots\,x_n\,\overline{x_1}\,\overline{x_2}\,\overline{x_3}\,\overline{x_4}\,, \quad \text{and so} \\
&\vdots \\
ICCR^{k-1}(x) &= x_k\,x_{k+1}\,\ldots\,x_n\,\overline{x_1}\,\overline{x_2}\,\ldots\,\overline{x_{k-1}} \\
ICCR^k(x) &= \overline{x_{k+1}}\,\overline{x_{k+2}}\,\ldots\,\overline{x_n}\,x_1\,\ldots\,x_k\,, \quad \text{for odd } k.
\end{aligned}$$

This gives

$$\begin{aligned}
ICCR^{n-1}(x) &= x_n\,\overline{x_1}\,\overline{x_2}\,\ldots\,\overline{x_{n-1}} \quad \text{and} \\
ICCR^n(x) &= x_1\,x_2\,\ldots\,x_n = x\,, \quad \text{as } n \text{ is odd.}
\end{aligned}$$

It could also occur, depending on x, that $ICCR^k(x) = x$ for $k < n$, so $|m_x|$ is at most n, as desired. ☐

Using the methods of the previous lemmas, it is also straightfoward to prove the following results, though we omit the proofs since we do not use these results in the following sections.

Lemma 37. *Let $x \in B_n$. If n is odd,*

$$\{y \ : \ y \overset{\text{CCR}}{\sim} x\} \ = \ \{a \ : \ a \overset{\text{ICCR}}{\sim} x\} \cup \{b \ : \ b \overset{\text{ICCR}}{\sim} CCR(x)\} \ .$$

For example, if $n = 3$, let $x = 000$. Then let $Y = \{y \ : \ y \overset{CCR}{\sim} x\} = \{000, 001, 011, 111, 110, 100\}$. Under the *ICCR*, the strings $x = 000$ and $y = CCR(000) = 001$ have the classes $A = \{a \ : \ a \overset{ICCR}{\sim} x\} = \{000, 110, 011\}$ and $B = \{b \ : \ b \overset{ICCR}{\sim} y\} = \{001, 100, 111\}$, and we have that $Y = A \cup B$.

Similar to our notation from Section 7.1, let $N'(n)$ be the number of ICCR necklaces.

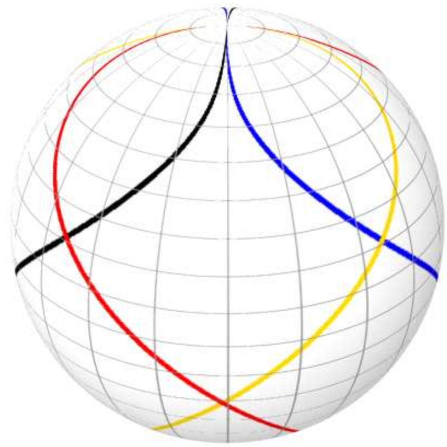
Corollary 12. *If n is odd, then*

$$N'(n) = 2N^*(n) = N(n) - \frac{1}{n}\sum_{2d|n} \phi(2d)2^{n/2d} \ .$$

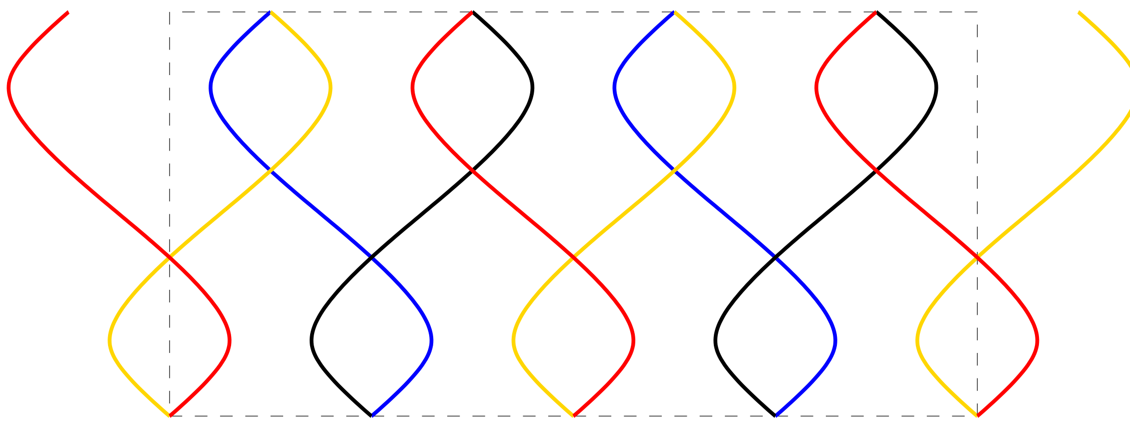
7.3 Motivating Examples

In this section we introduce some diagrams with curve-preserving rotary reflection symmetry that will illustrate the connection between such diagrams and the CCR and ICCR classes we have explored in previous sections. This section is joint work with Brett Stevens.

Consider the diagram shown in Figure 7.1, which we refer to as Diagram 7.1. Figure 7.1(a) shows the cylindrical projection; the actual area of the projection is shown inscribed in a rectangle, with extensions of the relevant curves across the vertical boundaries are included for clarity. Recall that in a cylindrical projection,



(a) Diagram on the sphere



(b) Cylindrical projection of the diagram

Figure 7.1: 4-Venn diagram with curve-preserving symmetry group S_8 on the sphere

the north pole maps to the horizontal line bounding the projection on the top side, so all of the points on the curves touching the top-most boundary are identified at a single point; similarly for the south pole.

There is a curve-preserving symmetry operation on Diagram 7.1, which is the rotary reflection $G_{\vec{v},\pi/4}$, where the axis of rotation specified by $\vec{v}$ is the vertical axis through the poles; the curve-preserving symmetry group of the diagram under this operation is the polydromic group S_8. Note that the diagram has other reflective symmetries across vertical planes, but in this chapter we are only interested in the polydromic symmetry groups and so we will ignore the other symmetries, if present.

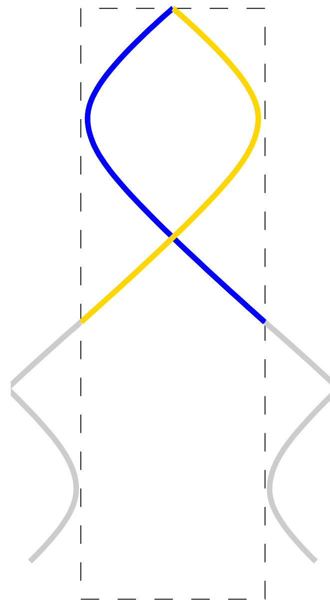


Figure 7.2: Fundamental domain for Diagram 7.1(b)

A fundamental domain, then, in Diagram 7.1 is given by a region covering 1/8th the surface of the sphere: for example, the area on one hemisphere bounded by two lines of longitude $\pi/4$ apart. Figure 7.2 shows a fundamental domain for the cylindrical projection of the diagram.

Composing this operation four times gives the oriented total symmetry $G_{\vec{v},\pi}$, which is the inversion symmetry: each curve maps onto itself but the interior and exterior of each curve have been swapped. This is since, if we consider one application of $G_{\vec{v},\pi/4}$ and consider a particular point p not on a curve, $G_{\vec{v},\pi/4}(p)$ will map p from the interior to the exterior (or vice versa) of exactly one curve, or exactly three (all but one) curves, depending on the location of p. Since the diagram is a Venn diagram, the next application of $G_{\vec{v},\pi/4}$ must move $G_{\vec{v},\pi/4}(p)$ from the interior or exterior (or vice versa) of a different one or three curves. Recall that a region can be defined by the set of interiors/exteriors represented by that region; If this process repeats a particular region before reaching the point in a region with exactly the complement set of interiors/exteriors of p, say after $G_{\vec{v},\pi/4}(p)^k$ with $k < 4$, then the diagram

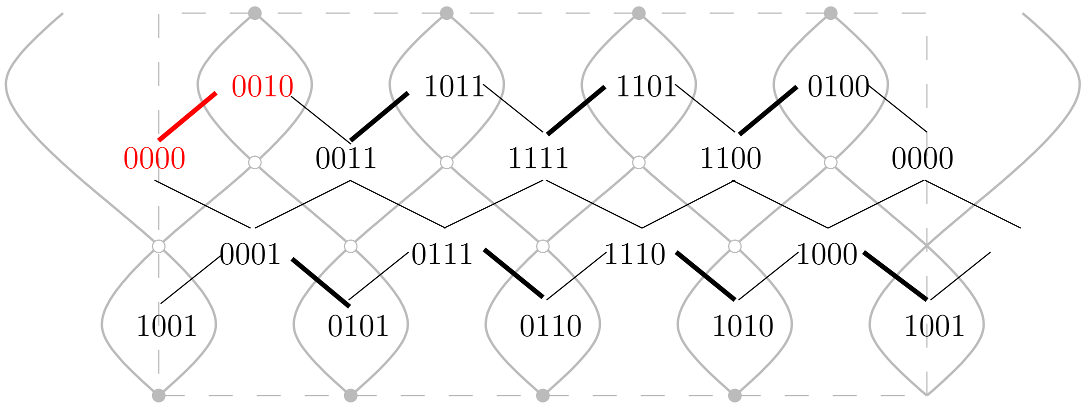


Figure 7.3: Dual of Diagram 7.1 illustrating symmetry group S_8

is not a Venn diagram, and if the complement region to p is reached after $k < 4$ applications of $G_{\vec{v},\pi/4}(p)$, then applying it $2k$ times would return to the complement of the complement—namely, the same region—before rotating all of the way around the sphere, and so again the diagram would not be Venn.

The operation $G_{\vec{v},\pi/4}$ on Diagram 7.1 maps a point on curve C_i where $1 \leq i \leq 4$, onto some other curve C_j, for any suitable labelling that induces a circular permutation on the curve labels. Applying $G_{\vec{v},\pi/4}$ eight times to return the diagram to the identity thus cycles each curve through each other curve twice before it returns to itself with the same orientation.

Now we examine the consequences of the above reasoning on the dual graph of the diagram, shown in Figure 7.3. A fundamental domain of the dual is shown in red; these two bitstrings generate the entire dual under the rotation. In this case each bitstring in the orbit of 0010 differs from the next by complementing three bits, corresponding to the three curves that are crossed when a point in the region labelled by a rotation of 0010 maps to the next. Each bitstring in the orbit of 0000 maps to the next by complementing one bit. Because the dual, as well as the diagram itself, must display the rotary reflection symmetry, for each copy of the dual to map onto the next, the dual is copied and flipped upside down as the bits are changed.

An interesting structure on the dual can be inferred by considering the cycles induced by the two labels. We have already noted that the curves can be arbitrarily labelled and each labelling will induce a circular permutation of the curve labels; the different curve numberings correspond to different indexing of the bitstrings labelling the regions. Of the 4! possible labellings, consider the one implicit in Figure 7.3. This labelling gives blue → 3, black → 2, red → 1, and yellow → 4, and this gives the following orbits for the two regions in the fundamental domain:

$$0000 \to 0001 \to 0011 \to 0111 \to 1111 \to 1110 \to 1100 \to 1000 \to$$

and

$$0010 \to 0101 \to 1011 \to 0110 \to 1101 \to 1010 \to 0100 \to 1001 \to .$$

These cycles are generated by the cycling complement shift register (CCR) function that we discussed earlier. Other curve labellings will produce the same cycles, but with the bits reordered, thus making the structure given by the CCR function less obvious.

Furthermore, the curve labelling of blue → 2, black → 1, red → 4, and yellow → 3 gives the following orbits for the two regions, with the region labelled 0010 in Figure 7.3 instead labelled as 0000:

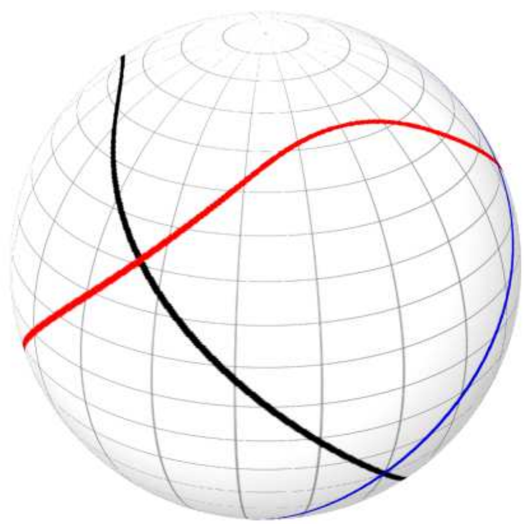
$$0000 \to 1110 \to 0011 \to 1000 \to 1111 \to 0001 \to 1100 \to 0111 \to$$

and

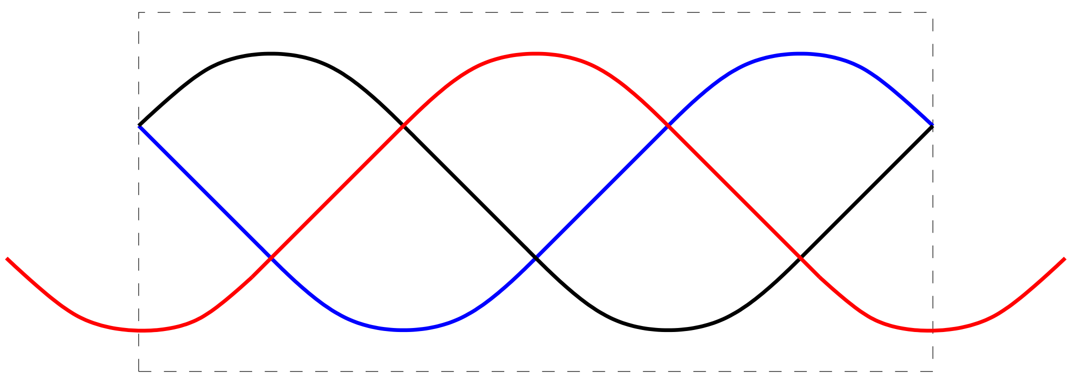
$$0010 \to 1010 \to 1011 \to 1001 \to 1101 \to 0101 \to 0100 \to 0110 \to .$$

These cycles are generated by the inverting complementing shift register (ICCR) function from Section 7.1.

As another example, consider the 3-Venn diagram in Figure 7.4. This diagram is further discussed in Section 8.2.1 in the context of its full antiprismatic dihedral



(a) Diagram on the sphere



(b) Cylindrical projection of the diagram

Figure 7.4: 3-Venn diagram with symmetry group S_6 on the sphere

symmetry group; we consider here that it also has the symmetry group S_{2n} under the rotary reflection $G_{\vec{v},\pi/n}$ where here $n = 3$, but it has the important structural difference that there are not, as in Figure 7.1, two vertices on the the axis of rotation through the poles. The dual is shown in Figure 7.5. Given the curve labelling red $\rightarrow$ 1, black $\rightarrow$ 2, and blue $\rightarrow$ 3, the dual shows how the region labels divide into the two orbits

$$000 \rightarrow 001 \rightarrow 011 \rightarrow 111 \rightarrow 110 \rightarrow 100 \rightarrow$$

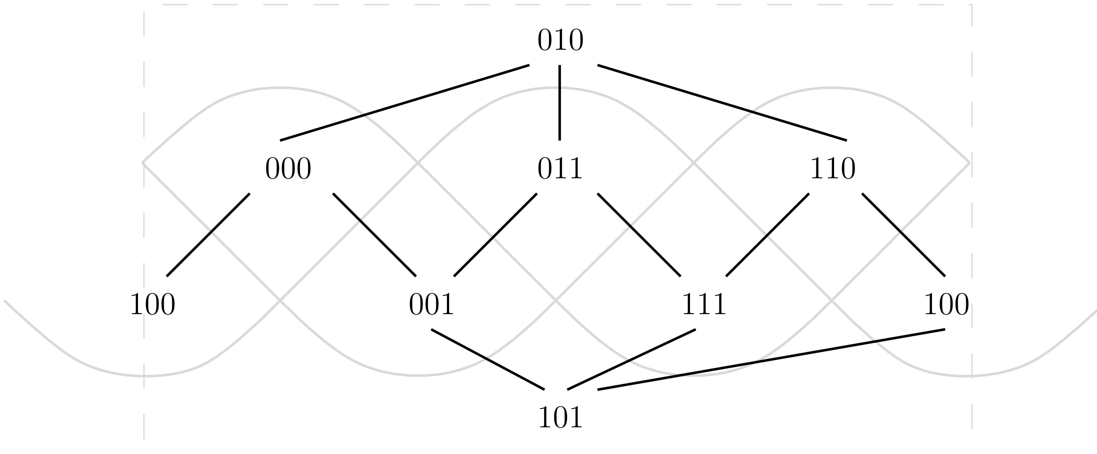


Figure 7.5: Dual of diagram in Figure 7.4

and

$$010 \to 101 \to ,$$

which, again, are generated by the CCR. Due to Lemma 35, it is easy to see that no matter what curve labelling is applied, the orbits of the region labels are not generable by the ICCR.

These observations are not coincidental, and we will see in future sections how the CCR and ICCR can be used to generate n-curve diagrams with similar symmetry groups, for many (but not all) different n.

7.4 Conditions for Rotary Reflection Symmetry in Venn Diagrams

Before discussing applying the CCR and ICCR directly to construct diagrams, in this section we discuss necessary conditions for an n-Venn diagram to have curve-preserving symmetry group S_{2n} on the sphere. For such a symmetry the isometry involved is $G_{\vec{v},\pi/n}$. This section is joint work with Frank Ruskey.

Fermat pseudoprimes are composite numbers n such that $\alpha^{n-1} \equiv 1 \pmod{n}$ for some constant base α. This lemma uses the Poulet numbers (also called Sarrus numbers), which are Fermat pseudoprimes of base 2 (*i.e* $\alpha = 2$). The five smallest Poulet numbers are 341, 561, 645, 1105, 1387 (see [128]), so in practical terms we need only worry about the other two cases in the following lemma for n when constructing n-Venn diagrams. Note that prime and Poulet numbers are odd for $n > 2$.

Lemma 38. *For an n-Venn diagram to have curve-preserving symmetry group S_{2n} on the sphere, with corresponding isometry $G_{\vec{v},\pi/n}$, n must be prime, a Poulet number, or a power of two.*

Proof. The single axis of rotation $\vec{v}$ can either pass through regions or through curves; each case is treated separately.

Case 1: *Axis of rotation passes through regions* : Then there are two regions, each containing a pole of rotation, that map onto each other under the symmetry. Then the labelling of the regions must be such that the polar regions belong to an orbit of size two. Similar to the proof of Theorem 1 (recently updated in [127]), all other regions must belong to orbits of size $2n$ for the rotation by π/n to map them onto themselves. Then $2n \setminus 2^n - 2$, or $n \setminus 2^{n-1} - 1$, thus $2^{n-1} \equiv 1 \pmod{n}$. The numbers for which this is true are either prime or Poulet numbers.

Case 2: *Axis of rotation passes through curves* : Consider a point x at either pole, and on a curve, and treat x as a vertex of the diagram (if it is on a single curve, we can treat it as a vertex of degree two). Recall from Section 4.2 that $G_{\vec{v},\pi/n}(G_{\vec{v},\pi/n}(x)) = R_{\vec{v},2\pi/n}(x)$ and so composing G maps x back onto itself, rotated by $2\pi/n$. Thus the degree of x must be divisible by n for $R_{\vec{v},2\pi/n}$ to be an isometry, and x must have rotational symmetry of order n. Since there are n curves, then x must be a vertex of degree either n or $2n$ (if x has degree n, then half of the curves pass through x, otherwise all curves pass through x).

Then there are n or $2n$ regions adjacent to each pole. Furthermore, similar to the previous case, all regions must belong to orbits of size $2n$, and thus $2n \setminus 2^n$, so $n \setminus 2^{n-1}$. Thus, n is a power of two since the prime factorization of 2^{n-1} consists solely of 2s.

□

Note that in the second case of the proof of Lemma 38, if the polar vertices are of degree $2n$ (with all n curves passing through each vertex) then each fundamental region will contain two edges adjacent to the poles and thus there will be two regions adjacent to the poles (if a fundamental region is chosen so that its borders follow edges), and each of these regions will belong to separate orbits, each of which has size $2n$.

7.4.1 Analogy between CCR and PCR Constructions

Thus, we have arrived at a way of understanding diagrams with this type of symmetry in terms of equivalence classes of binary strings, a method which we already know has been fruitful in formulating constructions for Venn diagrams on the plane (as in Chapter 3). The discussion is similar to our discussion of the GKS construction, building n-Venn diagrams with symmetry group C_n generated by applying the pure cycling register to necklace representatives making up a fundamental domain of the dual diagram, as discussed in Section 3.3.2.

Diagrams on the sphere resembling Figure 7.1 with the symmetry group S_{2n} have equivalence classes of strings that are of size $2n$ (with one class of size two in the case that poles pass through regions), and the CCR and ICCR give us a method of constructing such equivalence classes, just as the necklace function (PCR) gives equivalence classes that underlie the structure of planar symmetric Venn diagrams.

A summary of the analogous concepts is shown in Table 7.2. Recall that $N(n)$ is the number of PCR necklaces of B_n, from Section 7.1.

	GKS Construction	CCR/ICCR Construction
Symmetry group	C_n	S_{2n}
Order	n	$2n$
Restriction	n prime	n prime, 2^k, or a Poulet number
Number of labels in dual of fundamental domain	$N(n) - 2 = (2^n - 2)/n$	$2^{n-1}/n$ if n even, $(2^{n-1} - 1)/n + 1$ otherwise
Generating register	PCR	CCR/ICCR

Table 7.2: Characteristics of diagrams generated by PCR versus CCR/ICCR

In the following sections, we further examine the differences between the two types of diagrams, and exploit their similarities in order to build larger diagrams.

7.5 Constructing Symmetric Diagrams from the CCR Class

Recall from Section 3.3 the discussion of the GKS construction for planar symmetry n-Venn diagram for prime n. As we discussed in Section 3.3.2, choosing necklace representatives using block codes gives a necklace representative poset $\mathcal{R}_n$ that is then shown to have a chain decomposition with the chain cover property, and then this graph is treated as the dual of the fundamental domain of the final diagram. The dual is rotated around the point of rotation, rotating the node labels according to the necklace function each time, to give the dual of the final diagram.

In this section we discuss some differences between the GKS planar symmetric diagrams and ours. We will assume, since it is more generally applicable, that we will be using the CCR instead of the ICCR to partition the boolean lattice to build a dual of a fundamental domain. In Section 7.5.3 we will see why this appears to be justifiable.

7.5.1 Linking Edges and the Polar Face

One aspect of the GKS construction that needs consideration is that of linking the copies of the graph forming the dual graph of the fundamental domain of the diagram together. In the planar symmetric case, the point of rotation is contained in the central region, and so there is one vertex in the resulting dual of the diagram, labelled 1^n, located at that point, adjacent to which are the n vertices corresponding to all strings equivalent to the necklace $0\,1^{n-1}$. Similarly, the external region is represented in the dual as a point labelled 0^n, adjacent to which are all vertices corresponding to the necklace $0^{n-1}\,1$. Between these two extremal points the chains of the dual are embedded, and the resulting vertices in the final diagram are located at the same distance from the central point of rotation between the chains.

All of these properties are revised when we now consider the rotary reflective symmetric diagrams on the sphere.

Let us assume that some set of CCR necklace representatives can be found such that a chain decomposition embedding can be drawn with chain cover edges such that all curves in the resulting fundamental domain pass monotonically from one side to the other. We will need to use some extra edges to join all of the copies together, just as in the GKS construction the extra edges between the duals of fundamental domains and the 0^n and 1^n points creates the full, two-connected, dual of the entire diagram.

Definition. Let D be a diagram with some non-trivial symmetry, and let G be the dual graph of D, where G is formed from repeated copies of some subgraph forming the dual of a fundamental domain of D. A *dual domain linking edge* in G, or *linking edge*, is an edge that links copies of the dual of a fundamental domain, and thus crosses the boundary of the fundamental domain.

In the case that $n > 1$ is odd, then we have seen in the proof of Lemma 38 that the curves do not pass through the axes of rotation, and so building the dual of the fundamental domain of the diagram is similar to building the dual in the planar symmetric case: there are two "special" vertices, one at each pole, which are adjacent to k other vertices in the full dual, where k is a multiple of n. These two polar vertices map onto each other under the rotary reflection and so their labels must be the strings

$x = 0101\ldots10$ and $CCR(x) = 1010\ldots01$, which are the two strings present in the only CCR class of size 2 for odd n.

In the case of n a power of two, the proof of Lemma 38 shows that there must be two vertices of degree n or $2n$, one each located at the north and south poles (through the axis of rotation). Thus, there is no vertex in the dual of the fundamental domain located at these points. We can refer to a face in the dual through which a pole of rotation passes as a *polar face*. An *extremal vertex* in the dual is a vertex adjacent to the polar face. We must be careful to ensure that one of the extremal vertices, which will correspond to regions in the final diagram adjacent to these poles, is adjacent to its neighbouring copies in the dual of the final diagram, to ensure that the polar face in the resulting dual has degree n or $2n$, as we discussed in the proof of Lemma 38. This must be case for both poles, *i.e* at both the top and bottom of the dual of the fundamental domain. This will then ensure that every curve crosses the pole once, as there will be two edges per curve adjacent to the polar vertex.

To ensure this correct structure in the dual, it would suffice to build a dual of a fundamental domain with a single pendant extremal vertex that is then connected to both of its neighbouring vertices in the full dual, as in Figure 7.6. The extremal vertices are labelled a and z in the dual of the fundamental domain, and to make the polar faces degree $2n$, it is necessary for a to be adjacent to $CCR(z)$ and $CCR^{-1}(z)$, and similarly for z.

A second approach, which we have found more fruitful in generation programs, is the following. We can build a dual of a fundamental domain that has two connected extremal vertices adjacent to the polar face, and connect each of these vertices to their neighbours by two CCR shifts, as in Figure 7.7. These extremal vertices are labelled a and b at the north pole and z at the south pole in the Figure. In this configuration, the adjacent CCR shift of the entire fundamental domain puts $CCR(z)$ in a position that it is not extremal, so it has no special adjacency requirements except to ensure

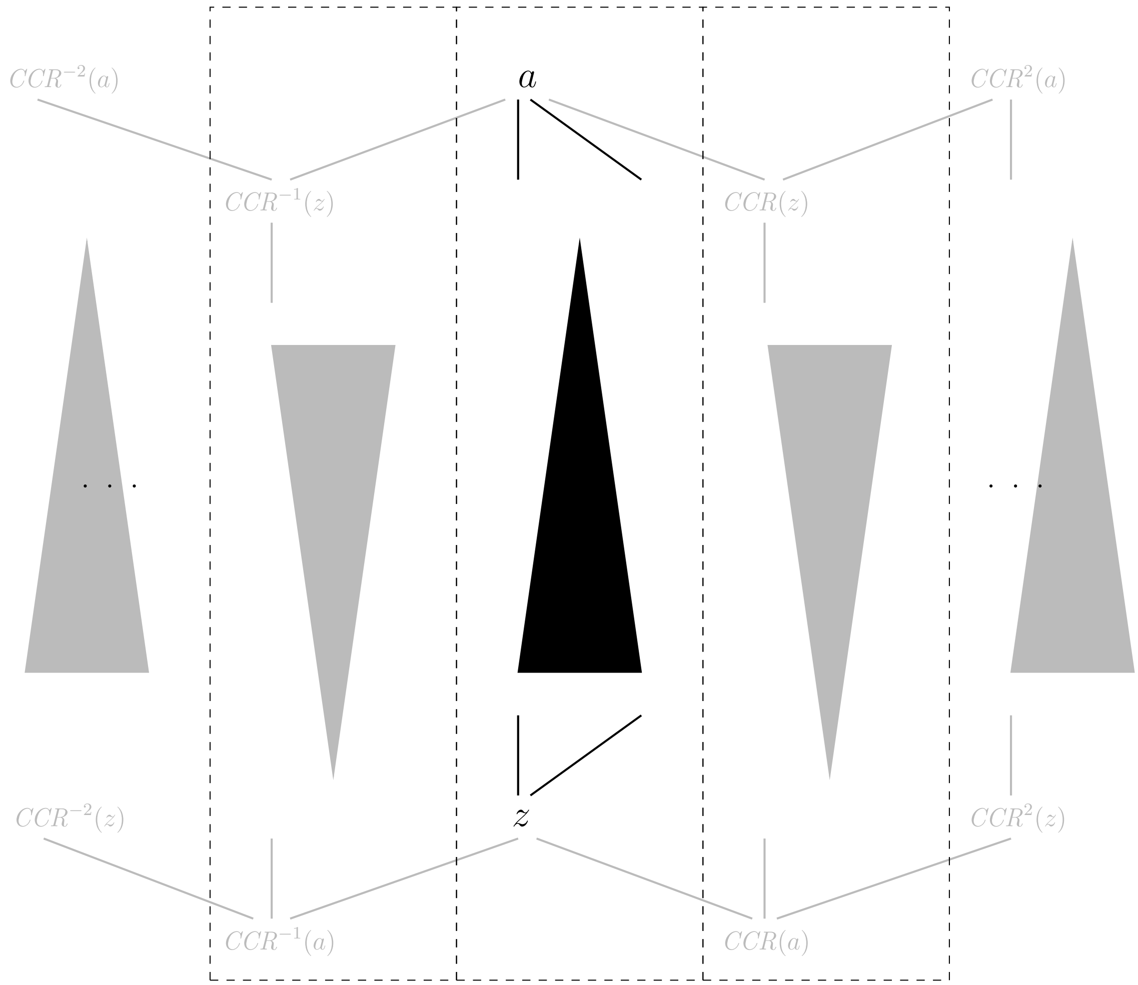


Figure 7.6: Structure of the dual of fundamental domains and copies, with adjacent copies joined by linking edges between a single extremal vertex in each dual of a fundamental domain; the black triangle represents the remainder of the dual between the extremal vertices

that it is two-connected to the rest of the graph, as usual. The 4-Venn diagram we have seen in Figure 7.1, with the dual in Figure 7.3, has this structure; in fact the entire structure of the dual is comprised of the two extremal vertices a and b and their connecting edge—every other edge adjacent to these two vertices is a linking edge.

7.5.2 Monotonicity Considerations

In the GKS construction for planar symmetric diagrams, the resulting diagrams are constructed to be monotone by the simple technique of building the dual from an

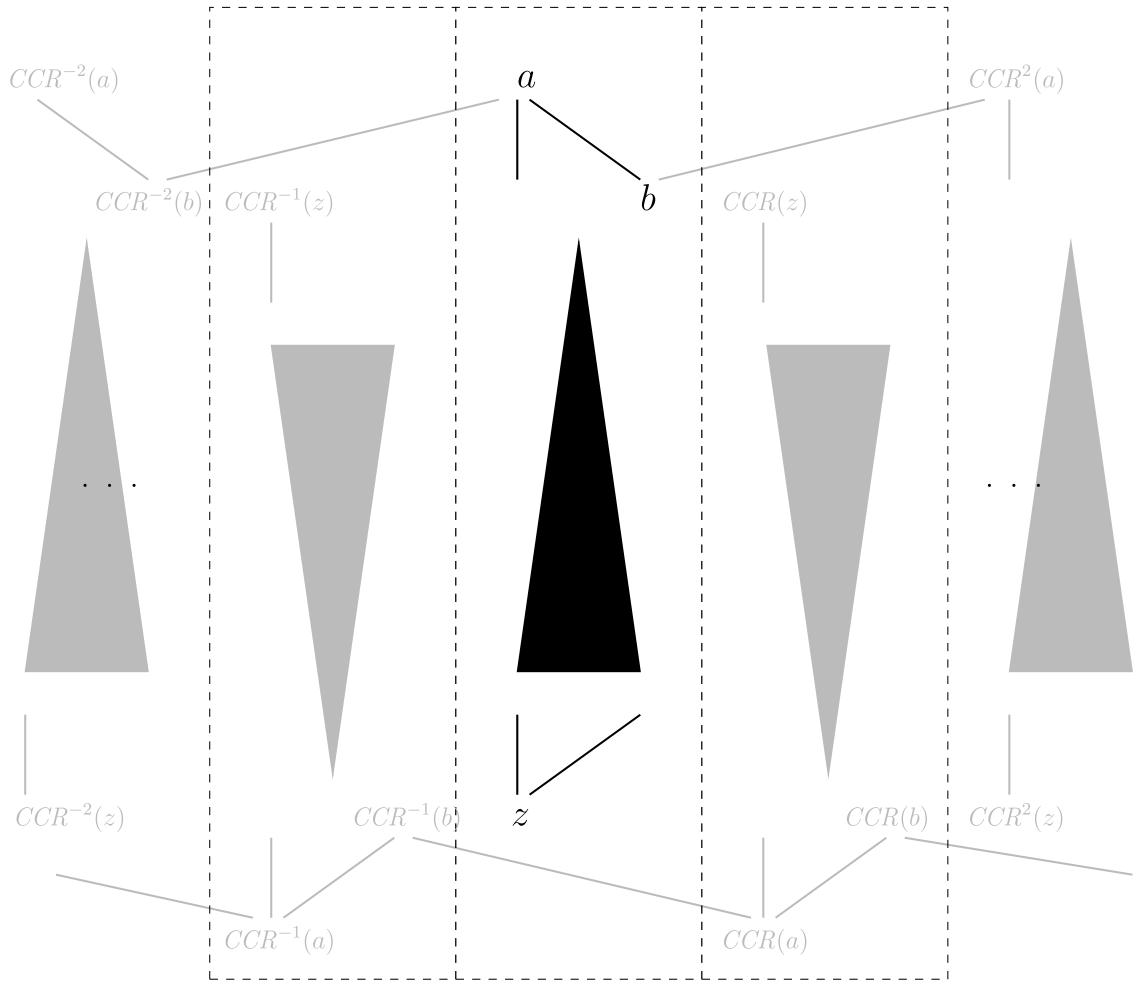


Figure 7.7: Structure of dual of fundamental domains and copies, with copies joined by linking edges between two extremal vertices to neighbours by two shifts

embedding of a chain decomposition for $\mathcal{R}_n$, the necklace representative poset found by considering block codes. This chain decomposition itself is monotone in the sense that each chain proceeds from the starter, of minimum weight, increasing in weight from each vertex to the next, to the terminator, of maximum weight. All vertices of rank k in the chain decomposition appear at the same horizontal position, with the lowest-rank vertices appearing higher (closer to the external face) and highest-rank vertices appearing lower (or closer to the internal face, in the resulting dual of the entire diagram). This property is preserved over the rotations of the dual of the fundamental domain, so all vertices of the same rank appear at the same distance from the internal and external faces throughout the dual. This technique ensures that the final diagram is drawn so that all curves proceed around the diagram such that a ray drawn from the point of rotation crosses each curve exactly once, and each curve forms a path around the central point of rotation that is connected into a loop and thus forms a simple Jordan curve, as desired.

On the sphere, in the case with n a power of two, each curve passes through each polar vertex once, and we must be more careful to ensure that in the resulting diagram built from the dual of the fundamental domain, each curve will link up with itself in such a way as to not form two or more components, but rather a single simple curve.

Since we are using the CCR (or ICCR) necklaces as opposed to the PCR necklaces, some of the monotonicity properties that come from using the PCR necklace are lost. First, given a string x, the weight $|x| \neq |CCR(x)| \neq |CCR^2(x)| \neq \ldots$ in general (though some strings, such as some $x = 01\ldots$, have the property that $|x| = |CCR^2(x)|$). This means that given a dual of a fundamental domain, if it is constructed by a chain decomposition of some poset, any properties of monotonicity will not be preserved in general as the weights of the labels in subsequent rotated copies of the dual of the fundamental domain will not be preserved.

However, this does not necessarily affect the ease of construction of the dual of the fundamental domain, since all that is necessary is to construct *one* dual of a fundamental domain that preserves the essential property of being monotone. We say that a chain decomposition embedding is *monotone-like* if, proceeding from extremal vertices in one direction to extremal vertices (if any) at the opposite side, bit i changes no more than once; this ensures that all curves in the resulting fundamental domain proceed across the dual from one side to another (*i.e* do not exit the fundamental domain from the same side they entered it on), or else they leave the fundamental domain from the top, or enter from the bottom (*i.e* cross the axis of rotation at either pole). (In the case with n odd, it is enough to ensure one fundamental domain can be built with curves proceeding monotonically across it.) If the dual of a fundamental domain can be constructed so that the correct number of curves in the resulting fundamental domain are adjacent to the polar vertex, then in the final diagram each curve will be adjacent to the polar vertex twice, ensuring that it passes through the polar vertex once, as desired.

7.5.3 Construction and Examples

In the GKS construction for planar symmetric n-Venn diagrams for prime n, one of the most significant tools used is the concept of the *block code*. The following construction gives us an analogous process to the GKS construction, excepting the choice of representatives using block codes.

Recall Lemma 35, which gives us an equivalence between the CCR and ICCR equivalence classes in the case that n is even. This allows us the freedom to choose to use the CCR or ICCR function to choose the necklace representatives for the dual of a symmetric diagram on the sphere.

The next theorem, like Lemmas 22 and 23 and others, is similar to Lemma 5.

Theorem 13. *Let n be a power of two. If there exists a set R_n of CCR necklace representatives or ICCR necklace representatives for B_n such that the following conditions hold:*

1. *the subposet $\mathcal{R}_n = (R_n, \leq)$ of $\mathcal{B}_n$ has a monotone-like chain decomposition embedding on the sphere $\mathcal{R}_n^*$ with the chain cover property,*

2. *dual domain linking edges can be added to $\mathcal{R}_n^*$ and its chain cover edges,*

then applying $G_{\vec{v},\pi/n}$ $2n-1$ times to $\mathcal{R}_n^*$ and its dual domain linking edges and chain cover edges, shifting node labels with the CCR or ICCR as appropriate, gives a dual graph D^* which is 2-connected and has both polar faces of degree n or $2n$, and D^* is *the dual graph of a spherical n-Venn diagram D that has polydromic curve-preserving symmetry group $S_C(D) = \mathsf{S}_{2n}$ under the operation of $G_{\vec{v},\pi/n}$.*

Proof. This proof is similar to the proof of Theorem 5 which we outlined in Section 3.3.2, using also the ideas from Hamburger's work [64, 65], with the added complications necessitated by the polar vertices in the resulting diagrams.

Given such a subposet $\mathcal{R}_n$ with a chain decomposition with the the chain cover property, chain cover edges, and dual domain linking edges, let $\mathcal{R}_n^*$ be the graph of the chain decomposition of R_n with chain cover edges embedded on the surface of the sphere. We will then rotate and copy it and take the dual to create the full n-Venn diagram. Let σ be the CCR shift function if the necklace representatives are for the CCR subposet, and the ICCR function if the necklace representatives are from the ICCR subposet.

We partition the surface of the sphere into $2n$ fundamental domains, separated by lines of longitude π/n radians apart. The graph $\mathcal{R}_n^*$ is embedded as a plane graph in one of the fundamental domains, as in Figure 7.8.

Extremal vertices adjacent to the north polar face are embedded uppermost (closest to the north pole) in $\mathcal{R}_n^*$, and extremal vertices adjacent to the south polar face embedded closest to the south pole. Now rotate the embedding $\mathcal{R}_n^*$ about the polar axis through $\pi i/n$ radians for each $1 \leq i \leq 2n-1$, and if i is odd, vertically reflect the embedding across the equator. In the ith rotation of the embedding, relabel each vertex x by $\sigma^i(x)$, and add copies of the dual domain linking edges between each copy of $\mathcal{R}_n^*$ (the locations of these edges are reflected for every other copy as well). Let D^* be the union of all such rotations and linking edges, along with the original embedding of $\mathcal{R}_n^*$.

The vertices of D^* are all subsets of $[n]$ corresponding with all binary strings, which each occur in exactly one necklace equivalence class as $\sigma^i(x)$ for some i and some $x \in R_n$, since the equivalence classes of x has size $2n$ by Lemma 11 and R_n

was chosen as a set of complete necklace representatives. Let D be the geometric dual of D^*; similar to the reasoning in the proofs of Lemmas 3 and 5, D forms an n-Venn diagram on the sphere. Curve C_i is made up of the edges crossing the i-edges in D^*, where an i-edge joins a vertex with label S with a vertex labelled $S \cup \{i\}$. At the boundary of a polar face in D^*, which is of degree n or $2n$ by the construction, there are two i-edges (or none), one joining two vertices a and b with label$(a) = S$ and label$(b) = S \cup \{i\}$ and the other joining $\sigma^n(a)$ and $\sigma^n(b)$. From Observation 6, label$(\sigma^n(a)) = \overline{S}$ and label$(\sigma^n(b)) = \overline{S \cup \{i\}}$, and so the edge between $\sigma^n(a)$ and $\sigma^n(b)$ is the other i-edge. Thus in D there are exactly zero or two i-edges incident to the polar vertex at each pole, so each curve C_i crosses the polar vertex at each pole once if at all.

Due to the monotone-like property of chain decomposition, each curve can be drawn so that it proceeds around the sphere such that a line of longitude drawn on a hemisphere spanning $\pi/2 > \phi > -\pi/2$ between poles, at any particular θ will cross the curve at most once (in fact it may not cross it at all if that line of longitude is near where a curve crosses a pole; for example, in Figure 7.9 a line of longitude between the poles drawn passing through the right-hand degree-14 vertex does not cross the yellow curve, since the yellow curve crosses the north pole in the fundamental domain shown). Thus each curve C_i is composed to two sections between the two points at which C_i cross the poles, and combined together these two sections give a simple cycle. This is similar to the argument in the proof of Lemma 3 that each curve forms a simple cycle.

Finally, we want to ensure that the resulting Venn diagram has unoriented curve-preserving symmetry group S_{2n}, by showing that the rotary reflection $G_{\vec{v},\pi/n}$, where $\vec{v}$ defines the axis of rotation passing through the poles, is an unoriented curve-preserving symmetry for the diagram D. By the construction it is clear that $G_{\vec{v},\pi/n}$ is an isometry for D^*, with curve labels permuted according to σ. The vertices and edges of D are constructed by the same procedure as in Lemma 3. The vertex corresponding to the north polar face is located at the position $\phi = \pi/2$ to coincide with the intersection of $\vec{v}$ with the sphere at the north pole, and it is identified within each subsequent copy of the dual of the fundamental domain, and similarly for the south polar face at position $\phi = -\pi/2$. To show that $G_{\vec{v},\pi/n}(C_i) = C_j$ for some $j \neq i$ we show that for each i-edge x, $G_{\vec{v},\pi/n}(x)$ gives some j-edge x'. The edge x is an i-edge because it crosses the edge (p, q) such that $label(p) = S$ for some set S and $label(q) = S \cup \{i\}$. Then the edge x' crosses the edge between $p' = G_{\vec{v},\pi/n}(p)$ and $q' = G_{\vec{v},\pi/n}(q)$. By the construction of D^*, we have $label(p') = \sigma(label(p))$ and $label(q') = \sigma(label(q))$, and so $label(p')$ and $label(q')$ differ by the application of σ to $\{i\}$, which is $i - 1 \bmod n$. This is true for all i-edges, so $G_{\vec{v},\pi/n}$ is a curve-preserving symmetry for each curve C_i, $1 \leq i \leq n$, and so it is an (unoriented) curve-preserving symmetry for D.

Since the order of the symmetry $G_{\vec{v},\pi/n}$ is $2n$, and the operation is a rotary reflection, consulting Table 2.2 gives $S_C(D) = \mathsf{S}_{2n}$. $\square$

Theorem 13 is illustrated for $n = 4$ by Figure 7.1, with the dual of the fundamental domain and linking edges shown in Figure 7.3.

Lemma 39. *There exists an* 8*-Venn diagram with polydromic curve-preserving symmetry group* $S_C = \mathsf{S}_{16}$ *on the sphere.*

Proof. The following series of figures establish Lemma 39. Figure 7.8 shows the dual of the fundamental domain for the 8-Venn diagram with a rotary reflection curve-preserving isometry giving symmetry group S_{16}. Using the proof of Theorem 13, this dual can be copied another 15 times, forming the full dual of the Venn diagram, and the dual taken to give the diagram on the sphere. The fundamental domain of the resulting diagram is shown in Figure 7.9, the resulting diagram is shown on the sphere in Figure 7.10, and as a stereographic projection in Figure 7.11. $\square$

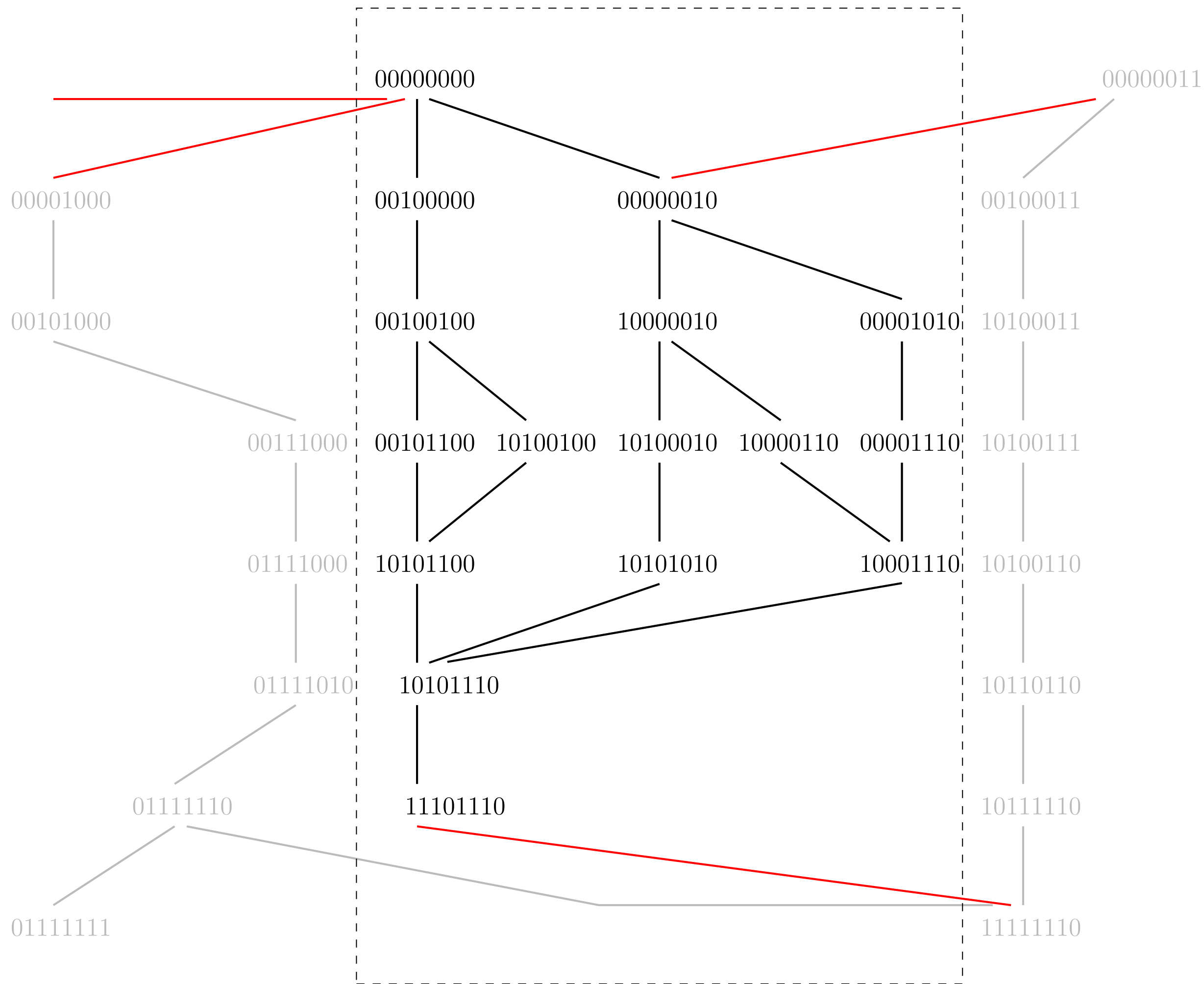


Figure 7.8: Dual of fundamental domain of 8-Venn diagram with rotary reflection isometry giving curve-preserving symmetry group S_{16}, with dual domain linking edges shown in red

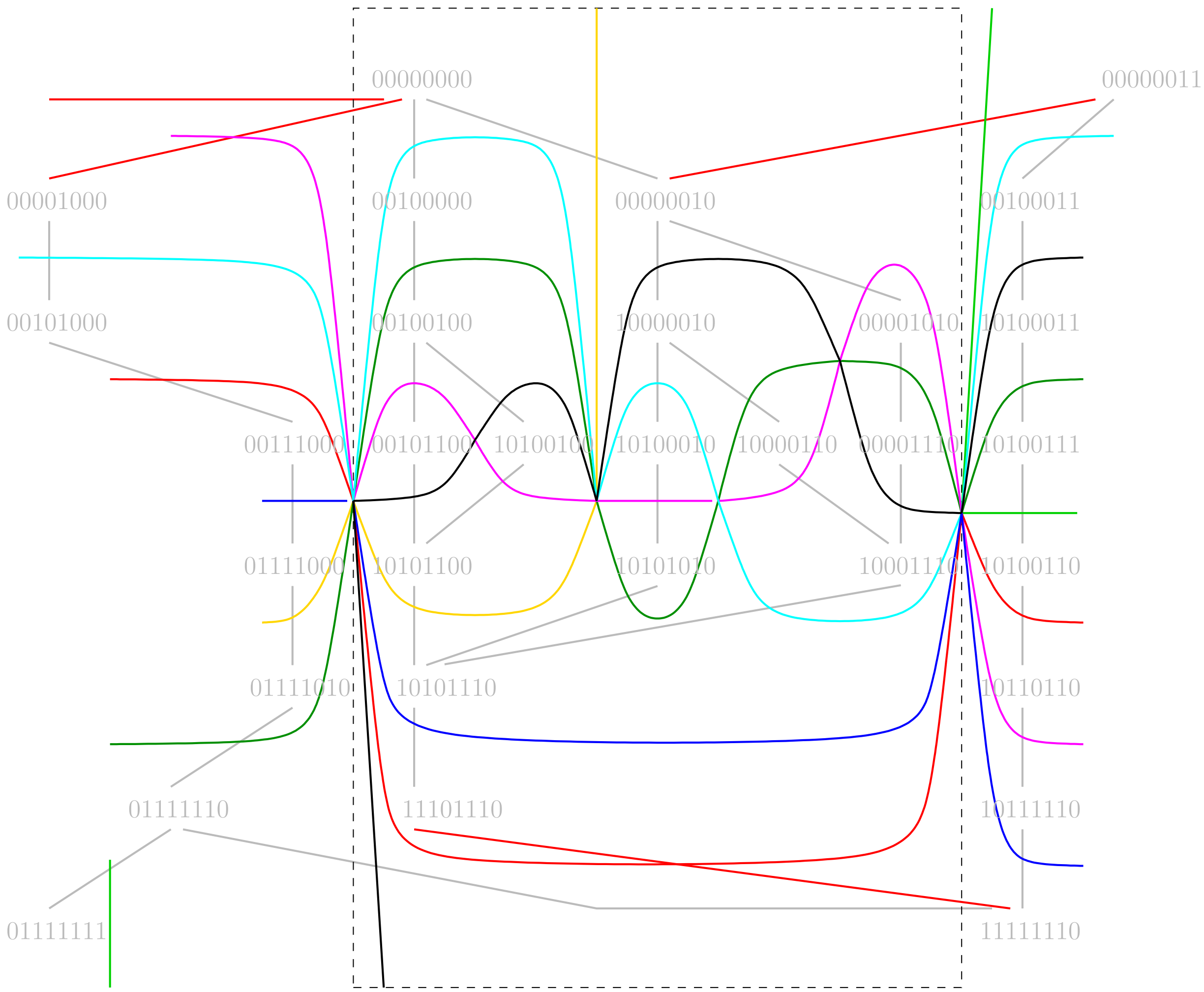


Figure 7.9: Fundamental domain of 8-Venn diagram with rotary reflection isometry giving curve-preserving symmetry group S_{16}

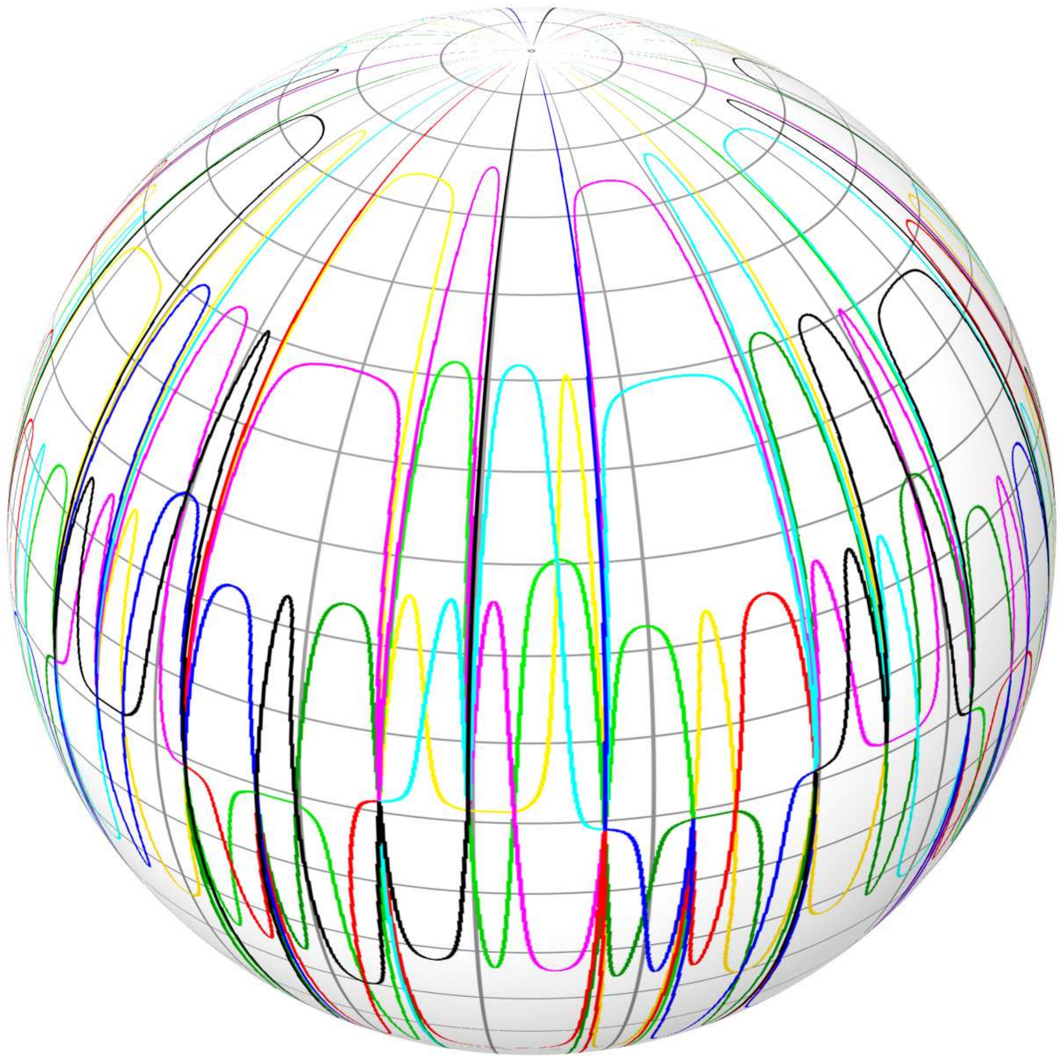


Figure 7.10: 8-Venn diagram on the sphere constructed from Figure 7.9

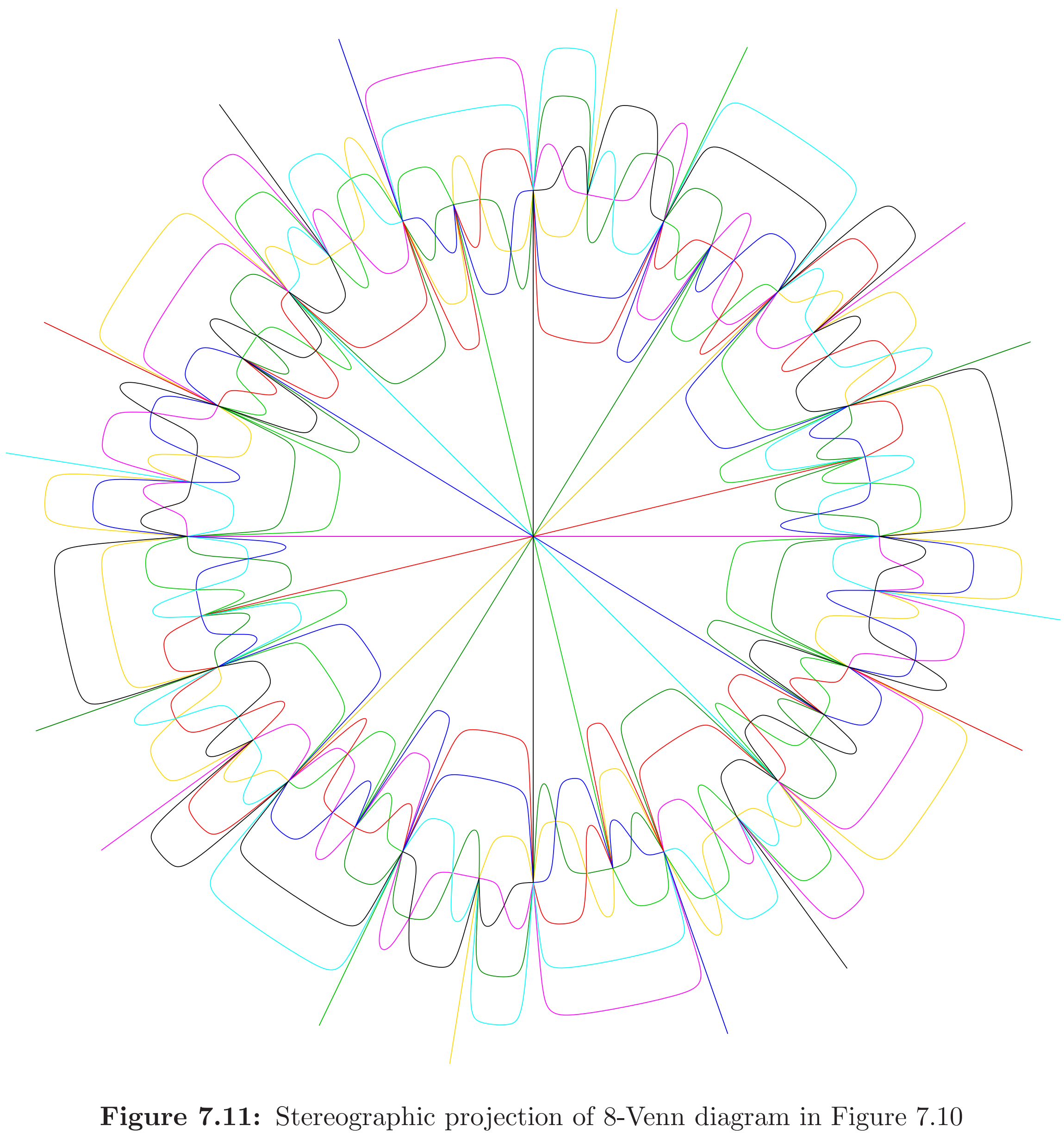


Figure 7.11: Stereographic projection of 8-Venn diagram in Figure 7.10

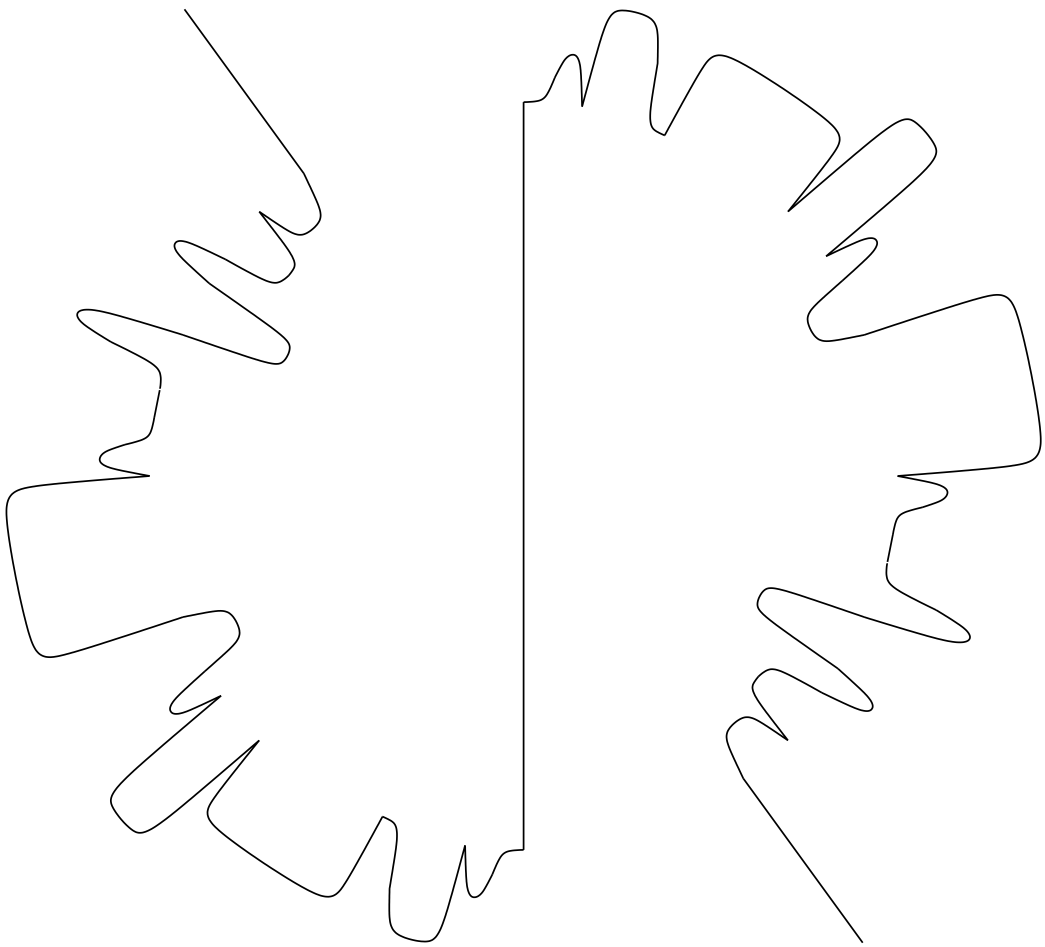


Figure 7.12: Single curve from 8-Venn diagram in Figure 7.11

For another example, the dual embedding of the 8-Venn diagram in Figure 7.13 is interesting in that every face, except the degree-16 face at each pole, is a 4-face, and so taking the dual gives a diagram which is simple except for the two necessary degree-16 vertices at the poles. Such a diagram we can call *nearly simple*, as it is as close as possible to simple that is achievable with this symmetry.

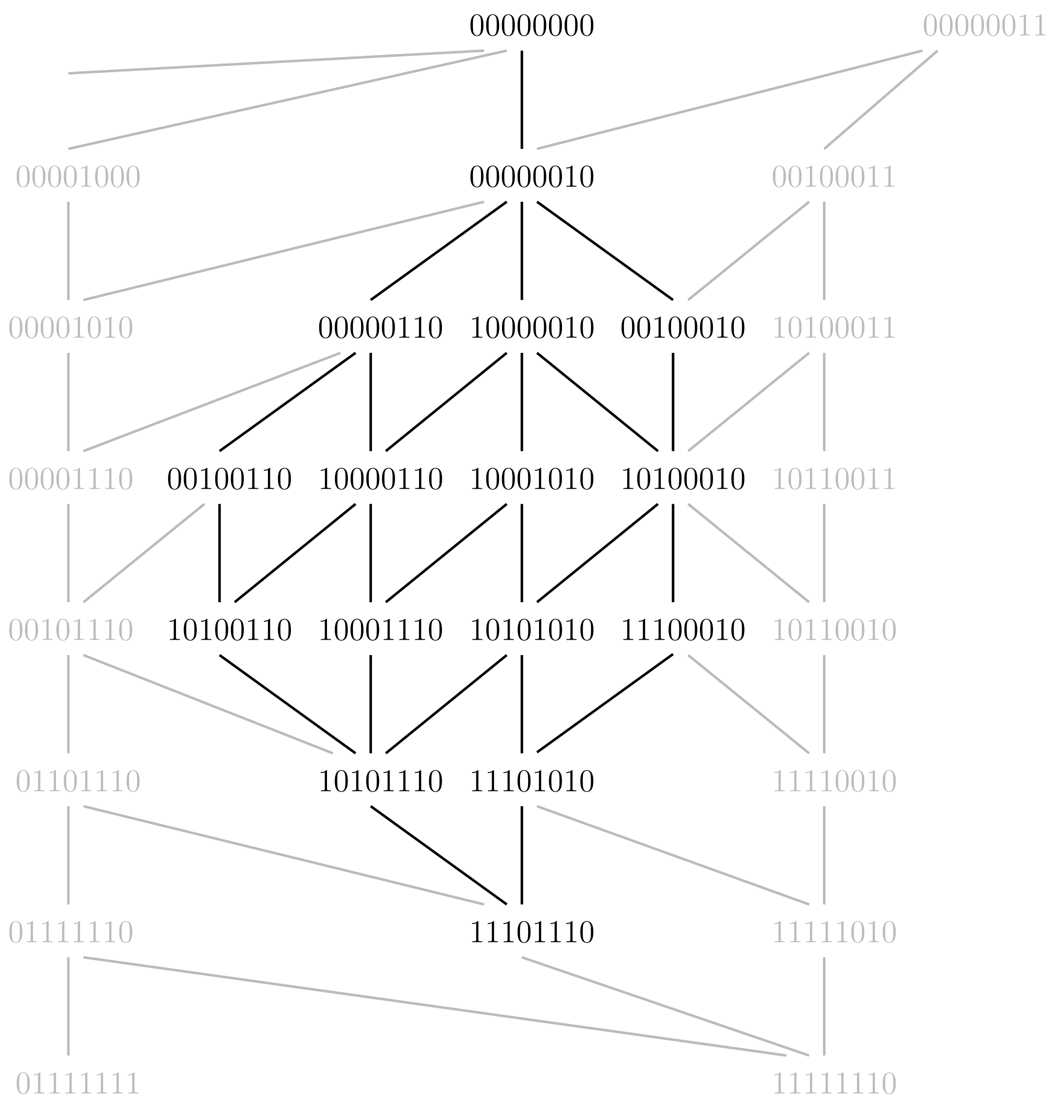


Figure 7.13: Dual of fundamental domain of nearly simple 8-Venn diagram with rotary reflection isometry giving curve-preserving symmetry group S_{16}

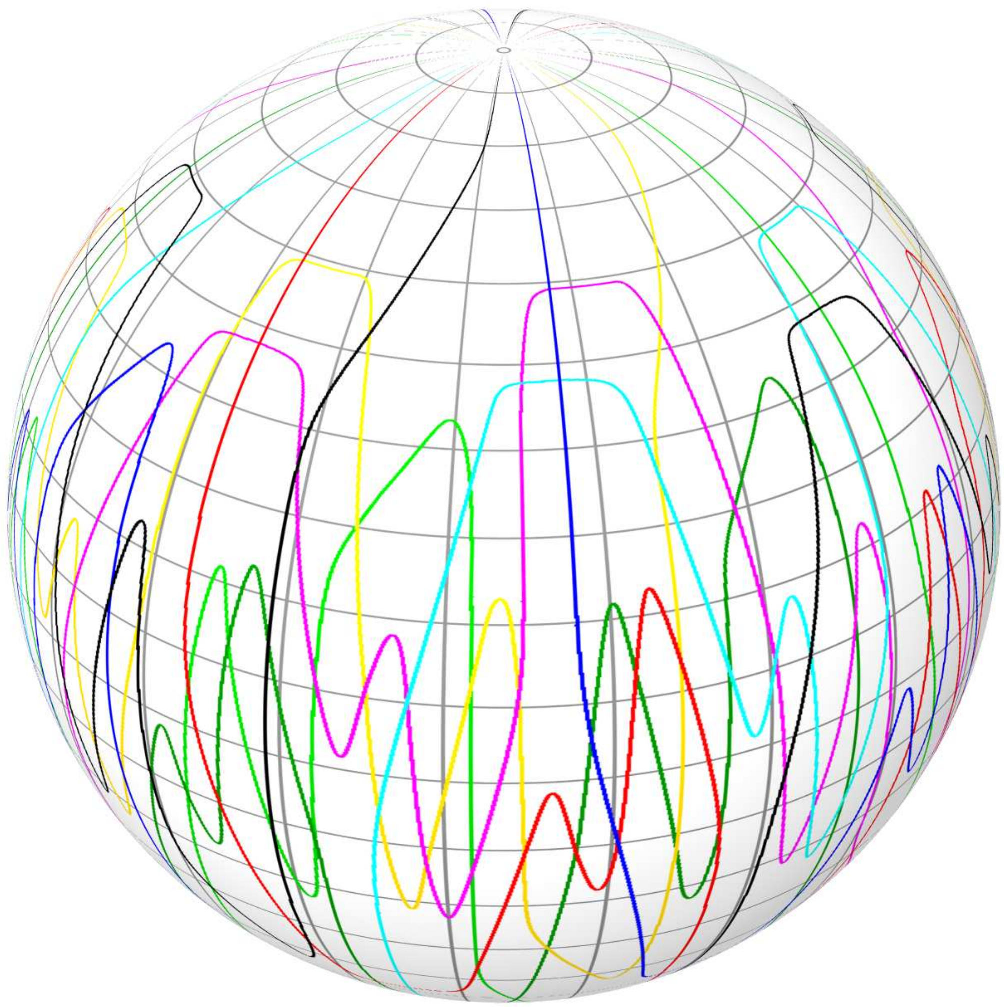


Figure 7.14: Nearly simple 8-Venn diagram on the sphere constructed from Figure 7.13

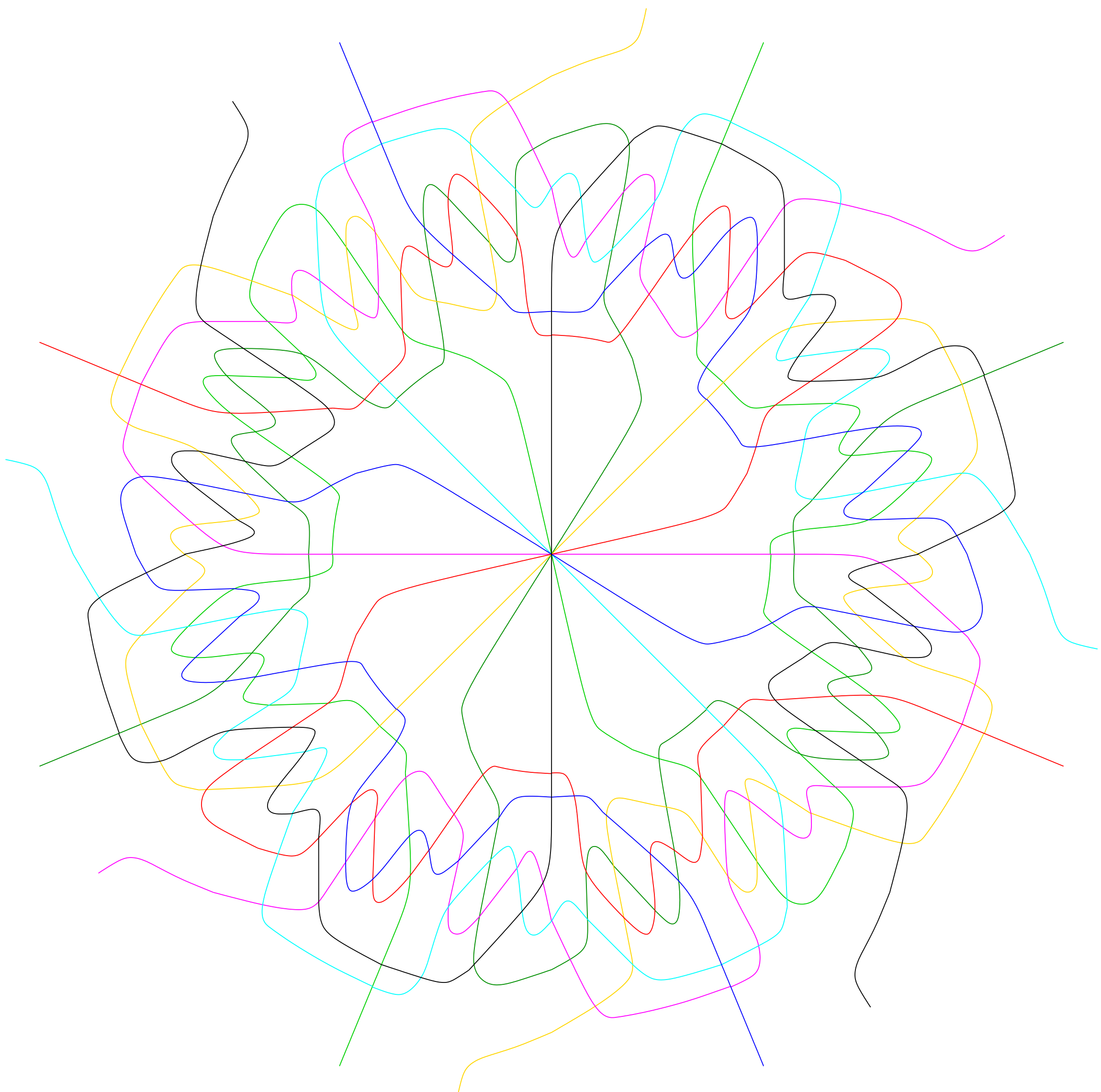


Figure 7.15: Stereographic projection of 8-Venn diagram in Figure 7.14

For the similar construction for n prime, as discussed in Section 7.5.2 the polar faces will be labelled by the equivalence class $\{0101\ldots0, 1010\ldots1\}$, and all other classes must be of size $2n$, which restricts us to at least the prime or Poulet numbers.

Let x be the string $(01)^{\frac{n-1}{2}}0$, so $CCR(x) = (10)^{\frac{n-1}{2}}1$. These are the strings that must be the labels of the polar vertices in the dual graph construction. Let $B_n^* = B_n - \{x,\ CCR(x)\}$.

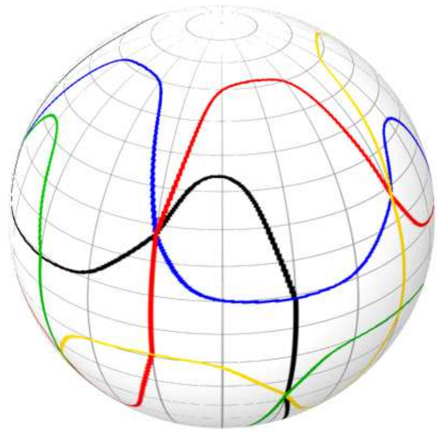
Theorem 14. *Let n be prime or a Poulet number. If there exists a set R_n^* of CCR necklace representatives for B_n^* such that the following conditions hold:*

1. *the subposet $\mathcal{R}_n^* = (R_n^*, \leq)$ of $\mathcal{B}_n^*$ has a monotone-like chain decomposition with the chain cover property,*
2. *$\mathcal{R}_n^*$ has a vertex adjacent to x or $CCR(x)$*
3. *dual domain linking edges can be added to the embedding $\mathcal{R}_n^*$ and its chain cover edges,*

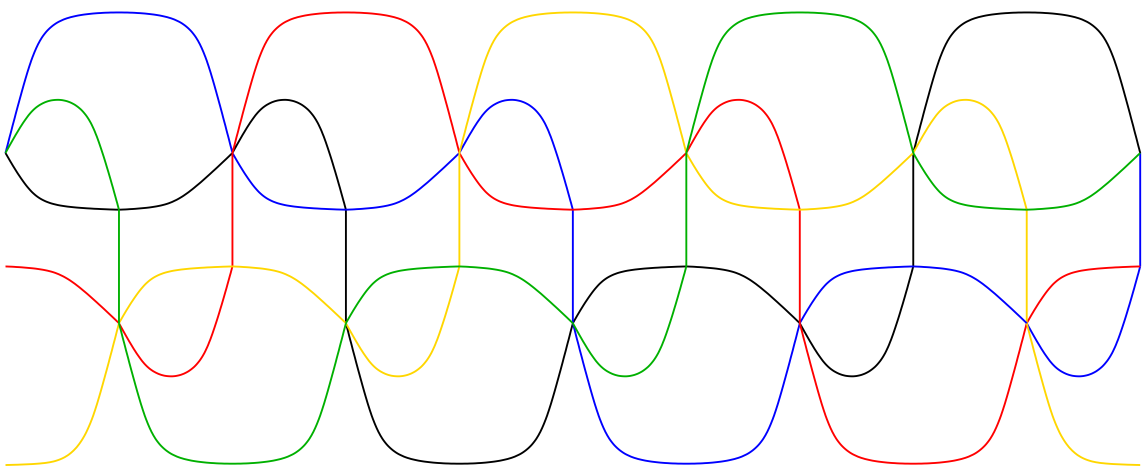
then applying $G_{\vec{v},\pi/n}$ $2n-1$ times to $\mathcal{R}_n^*$ and its dual domain linking edges and chain cover edges, shifting node labels using the CCR, gives a dual graph D^* which is 2-connected and has vertices adjacent to $\{x,\ CCR(x)\}$, and then D^* is the dual graph of a spherical n-Venn diagram D with polydromic curve-preserving symmetry group $S_C(D) = \mathsf{S}_{2n}$.

Proof. The proof is essentially the same as that for Theorem 13, without the added complications of having vertices at the poles, and thus the proof is more similar to the GKS Theorem 5. □

Figure 7.16 shows a 5-Venn diagram with polydromic symmetry group S_{10} that is generable by the CCR. The full dual is shown in Figure 7.18. Searches by hand have revealed no other 5-Venn diagrams on the sphere generable by the CCR, and so we believe this diagram is unique. Figure 7.17 shows the stereographic projection of the diagram, realizing the planar symmetry group C_5, which is isomorphic to a subgroup of the spherical symmetry group S_{10}.



(a) Diagram on the sphere



(b) Cylindrical projection of the diagram

Figure 7.16: 5-Venn diagram with rotary reflective isometry giving curve-preserving symmetry group S_{10} generable by the CCR

Similar to the 5-Venn diagram generable by the CCR, we have also found a small number of 7-Venn diagrams generable by the CCR; the dual of the fundamental domains have $(2^7 - 2)/14 = 9$ CCR necklace representatives. One example is shown in Figures 7.19 to 7.22. The stereographic projection of this diagram is shown in Figure 7.21; since the cyclic group C_n is isomorphic to a subgroup of the group S_{2n}, this diagram exhibits the cyclic group C_7 and thus is planar symmetric.

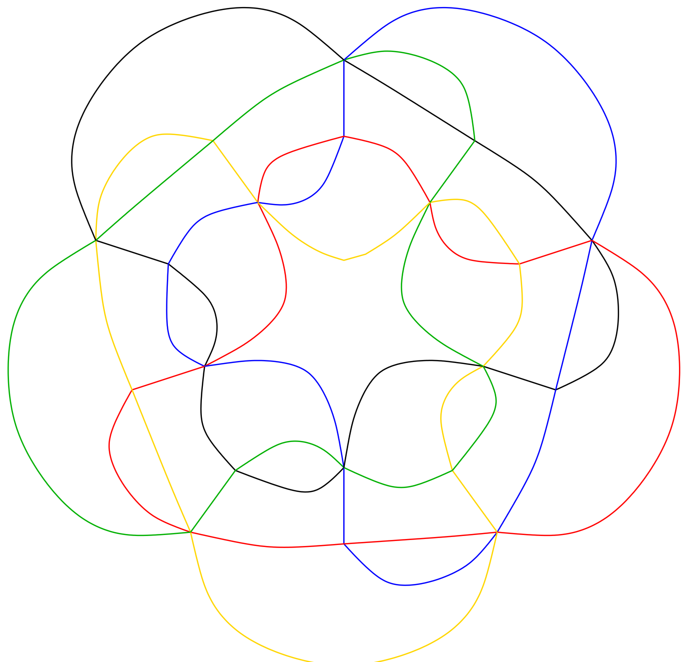


Figure 7.17: Stereographic projection of 5-Venn diagram in Figure 7.16, realizing planar symmetry group C_5

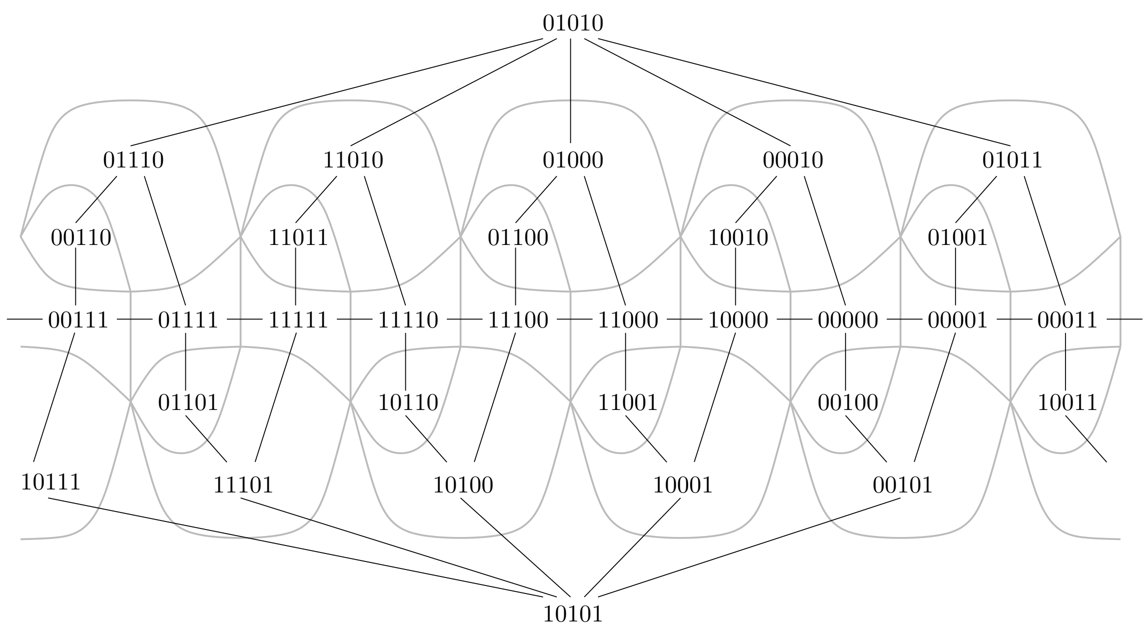


Figure 7.18: Dual graph of 5-Venn diagram in Figure 7.16

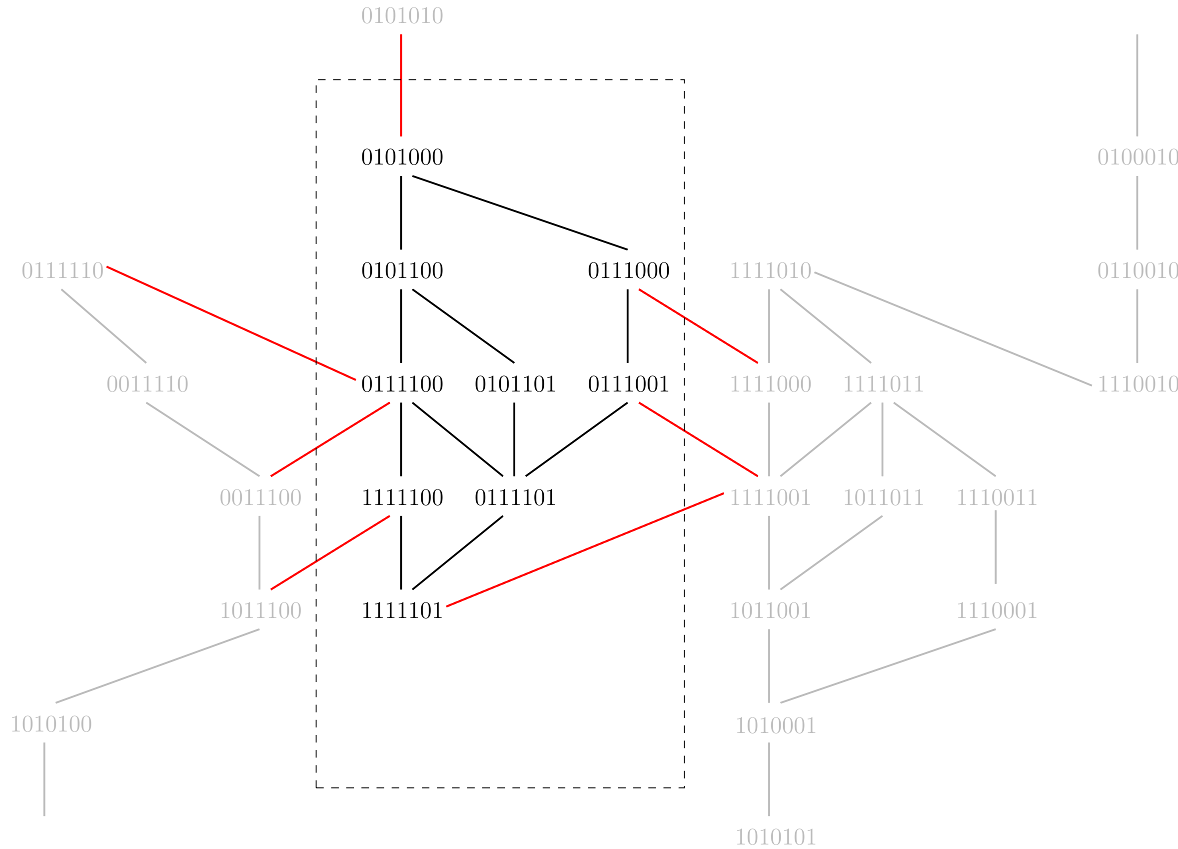


Figure 7.19: Dual of fundamental domain of 7-Venn diagram with rotary reflection isometry giving curve-preserving symmetry group S_{14}, with dual domain linking edges shown in red

The 7- and 8-curve diagrams in Figures 7.10 to 7.15 and 7.19 to 7.22 were discovered not by hand but by computer search over the dual graph. The backtracking technique is very similar to that used to search for symmetric n-Venn diagrams for prime n, symmetric independent families, symmetric near-Venn diagrams, and pseudo-symmetric n-Venn diagrams for n a prime power, by the authors of [107]. See [131] for more details on the backtracking code and data structures involved. Our computer search found many 8-Venn diagrams up to isomorphism with the polydromic curve-preserving symmetry group S_{16}, and their frequency, along with the diagrams we have found for $n = 4$, 5, and 7, leads us to the following conjecture.

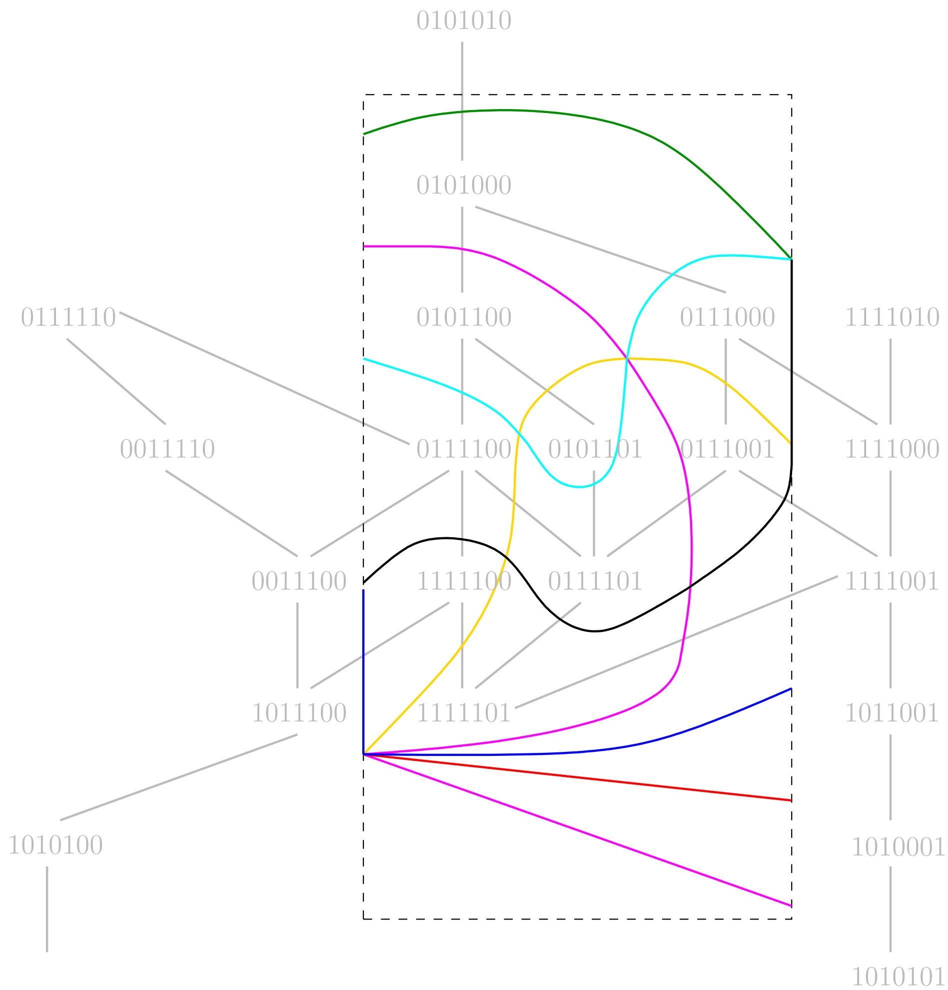


Figure 7.20: Fundamental domain of 7-Venn diagram with rotary reflection isometry giving curve-preserving symmetry group S_{14}

Conjecture 5. *For all n prime, a Poulet number, or $n = 2^k$ for $k \geq 1$, there exists an n-Venn diagram D on the sphere realizing polydromic curve-preserving symmetry group $S_C(D) \cong \mathsf{S}_{2n}$.*

For the case where $k = 1$, a 2-Venn diagram with polydromic symmetry group S_4 on the sphere can be formed from two great circles given by $\theta = \{0, \pi\}$ and $\theta = \pm\pi/2$; these curves form orthogonal great circles intersecting at the north and south poles, as in Figure 7.23. This diagram has three orthogonal reflective total symmetries, giving it the total symmetry group D_{2h} of order 8. For this section, our interest is

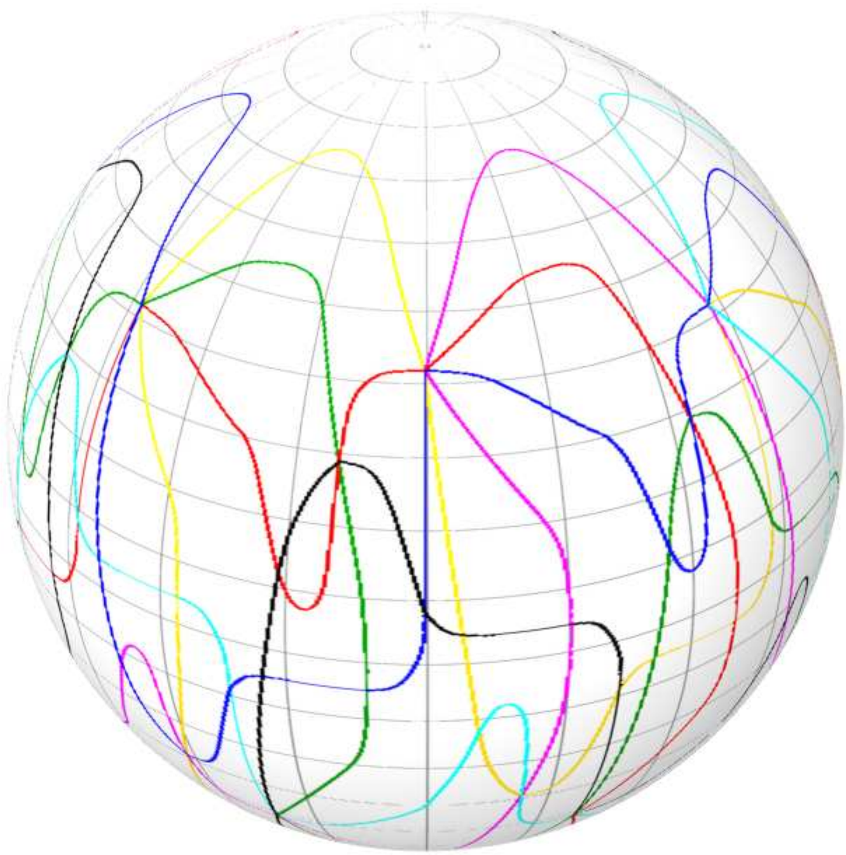


Figure 7.21: 7-Venn diagram on the sphere with curve-preserving symmetry group S_{14} constructed from Figure 7.20

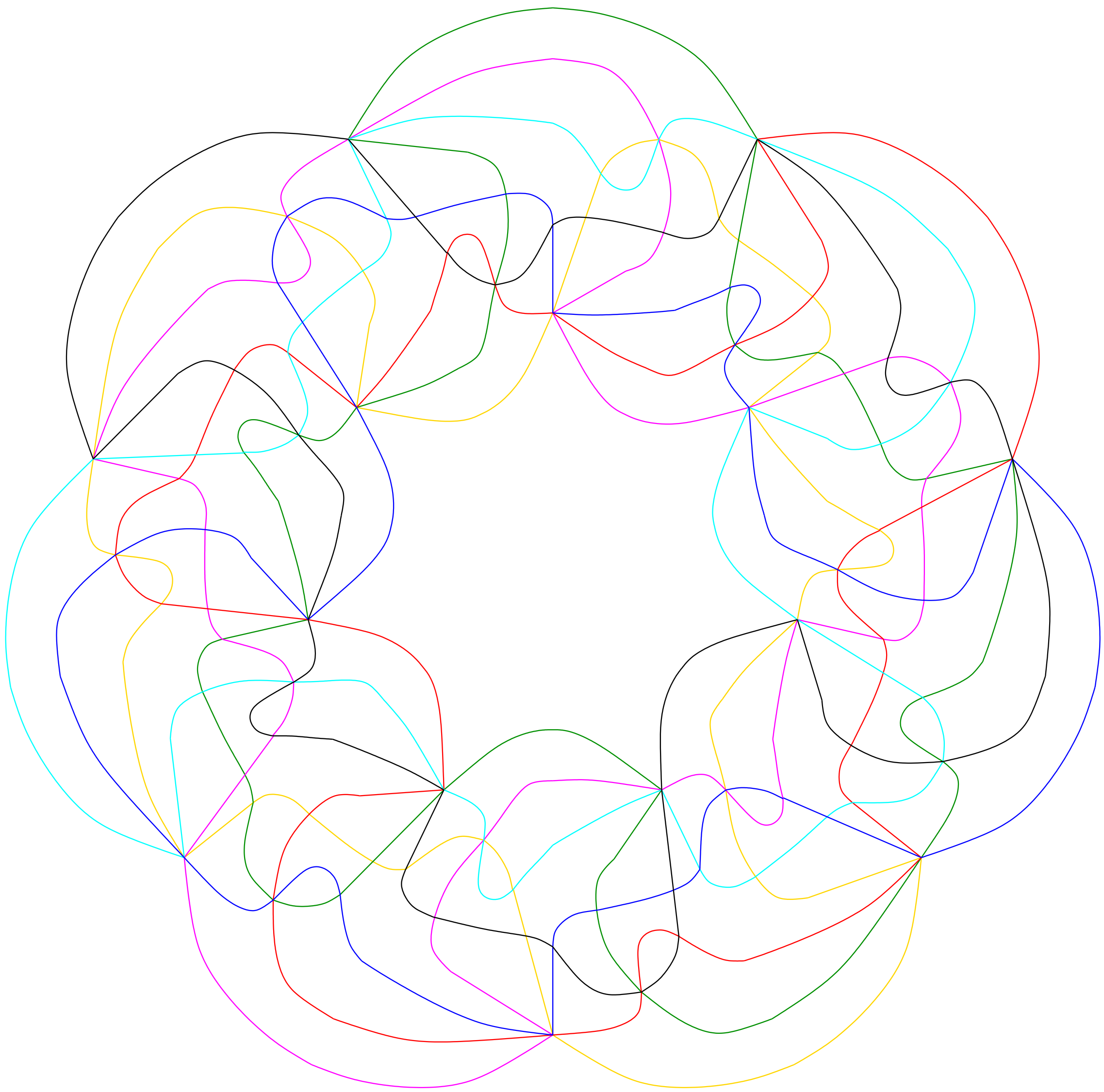


Figure 7.22: Stereographic projection of 7-Venn diagram constructed from Figure 7.20, exhibiting planar symmetry group C_7

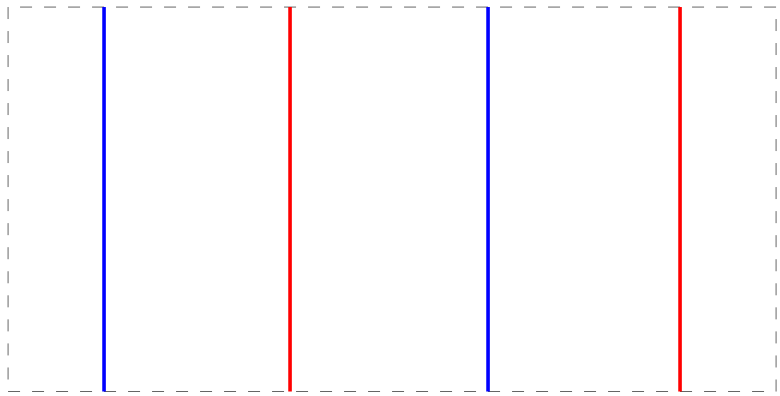


Figure 7.23: Cylindrical projection of 2-Venn diagram with total symmetry group S_4

in the polydromic symmetry groups, which are a subgroup of the prismatic groups, and this diagram has the curve-preserving polydromic group D_{4h} of order 16, which contains as a subgroup the group S_4 of order 4.

The next possibility, of 16-Venn diagrams on the sphere with symmetry group S_{32}, is far too large to construct by hand or by computer search using present techniques; the dual of a fundamental domain will contain $2^{16}/32 = 2048$ vertices.

Lemma 40. *Diagram 7.1 is, up to isomorphism, the only* 4*-Venn diagram with polydromic curve-preserving spherical symmetry group* S_8*.*

Proof. In any four-curve Venn diagram, all vertices must be either degree four, six, or eight, so we proceed by considering the possible vertex distributions. We know from the proof of Lemma 38 that the axis of rotation cannot pass through regions, so the only possibility for the poles is that the axis of rotation passes through two vertices, each of degree four or eight, and all other vertices will belong to orbits of size eight. Then the number of vertices is equal to $2 (\bmod 8)$. Since the number of faces is equal to 16, Euler's formula gives us $v - e + f = 2$ and thus $v - e + 14 = 0$. Let d_x be the number of vertices in the diagram with degree x. By summing vertex degrees, where the maximum vertex degree is eight, and substituting into Euler's formula we have $d_4 + d_6 + d_8 - (4d_4 + 6d_6 + 8d_8)/2 = -14$ and $d_4 + 2d_6 + 3d_8 = 14$. The following table shows the only possibilities that satisfy this equation with the total vertex counts equal to $v = 2 \pmod 8$:

d_4	d_6	d_8	$\lvert V\rvert$
6	4	0	10
8	0	2	10

Since the axis of rotation passes through two vertices, which will map onto each other under the rotary reflection, one of the vertex counts must be 2 mod 8 and the other 0 mod 8. This only leaves $d_4 = 8$ and $d_8 = 2$ as an option. All four curves must pass through each degree 8 vertex, at each pole, and a fundamental domain for the order-eight symmetry group S_8 must contain exactly one degree 4 vertex. By trying all configurations it is simple but somewhat tedious to verify that the only possible configuration for this vertex is to have two edges joining it to a degree 8 vertex at the pole and two edges connected to its neighbours that it maps onto under the rotary reflection, which gives the edge configuration of Diagram 7.1.

Finally, the curves can only be coloured in the configuration in Diagram 7.1. This is because there is no choice for curve colouring when labeling a degree-four vertex, the only choice is at the two degree-eight vertices, and it is again simple to verify that no other colouring works except that a given curve passes straight through the degree-eight vertex (*i.e* there are three intervening edges between the two edges of a given colour incident on a degree-eight vertex). □

We observed in Section 7.3 that the dual graph of Figure 7.1 can be labelled in one way to give the CCR equivalence classes but in another way to give the ICCR equivalence classes. Since, by Lemma 35 the equivalence classes are the same, it appears that the following is true in general.

Conjecture 6. *Let n be a power of two. A chain decomposition of the poset of CCR necklace representatives that gives an n-Venn diagram on the sphere with polydromic* *curve-preserving symmetry group* S_{2n} *according to the construction of Theorem 13, can be transformed into one using ICCR necklace representatives by applying a set of bit complementation operations and reordering bits to all labels, and vice versa.*

Recall our introduction of the boolean lattice in Section 2.3.1. It is interesting to note that the set of operations on bitstrings of length n of applying a set of bit complementation operations and reordering bits is exactly the set of functions that are automorphisms of the boolean lattice of order n.

7.6 Sufficiency Questions

In this section we step back to consider questions of sufficiency, namely when it is possible to use shift registers at all to construct diagrams, Venn or otherwise, with symmetry group S_{2n}. While it may seem plausible that *any* n-curve diagram (not necessary Venn) on the sphere with polydromic curve-preserving symmetry group S_{2n} has a dual that is generated by applying the complementing cycling register to an appropriate fundamental domain, we already know that in fact this is not the case, due to the necessary conditions from Lemma 38.

Figure 7.24 shows a 6-curve (non-Venn) diagram built using a construction that has two vertices of degree 12, one at each pole. Since n is even but is not a power of two, there are equivalence classes under $\overset{CCR}{\sim}$ (and $\overset{ICCR}{\sim}$) with cardinality smaller than $2n$. For example, the class

$$001100 \rightarrow 011001 \rightarrow 110011 \rightarrow 100110 \rightarrow \;,$$

cannot label any cycle of regions in the diagram as they all belong to orbits of size $2n$.

The construction in Figure 7.24 can easily be extended to any n; however if n satisfies Lemma 38 then these figures can be labelled with the CCR, simply by choosing appropriate curve labellings.

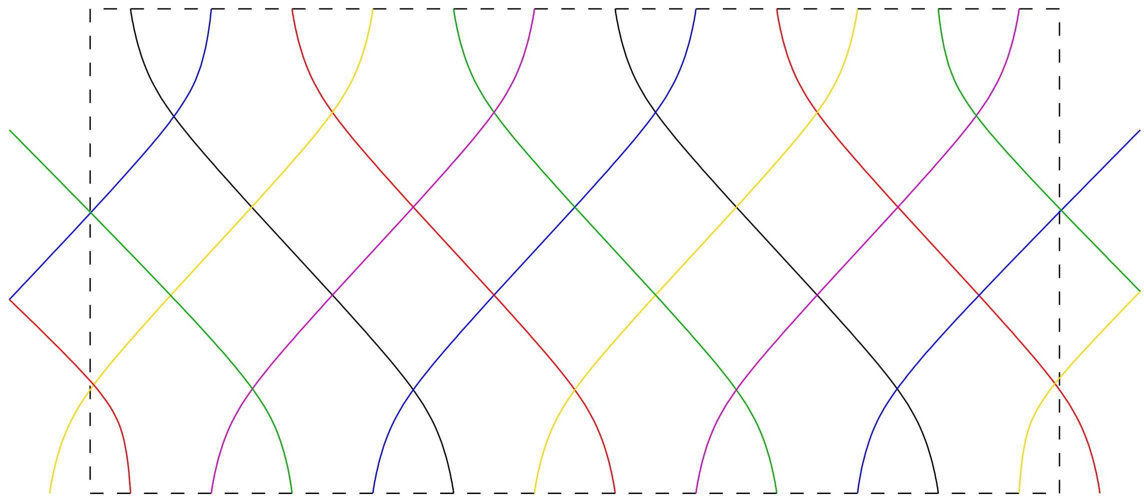


Figure 7.24: A 6-curve non-Venn diagram with rotary reflective isometry giving symmetry group S_{12} not generable by the CCR or ICCR

We have observed that all of the n-curve diagrams generated under the CCR/ICCR are n-Venn diagrams, and it appears difficult to construct by hand an n-Venn diagram for small n (only $n \leq 7$ is realistic to attempt by hand) that does not use the CCR. Thus we tentatively offer the conjecture:

Conjecture 7. *Let n be a power of two, prime, or a Poulet number. Any n-curve diagram (not necessarily Venn) with polydromic curve-preserving symmetry group S_{2n} under a rotary reflection isometry on the sphere, is generable by constructing the dual of a fundamental domain from representatives of the CCR or ICCR equivalence classes.*

7.6.1 Open Questions

The previous section has made some conjectures as to which diagrams can be constructed using the CCR and ICCR equivalence classes. The primary open question is: given any n a power of two, prime, or a Poulet number, can a chain decomposition be constructed for $\mathcal{B}_n$ such that it can be embedded and linked to form the dual of a fundamental domain for a full Venn diagram with curve-preserving symmetry group S_{2n} on the sphere, using the CCR or ICCR equivalence classes? The methods used in the planar case, for the GKS construction, do not seem to work when applied to the CCR/ICCR case. There could also be further restrictions on n that are unknown; thus it is worthwhile to investigate whether n can be restricted further from the necessary conditions given in Lemma 38. Finally, other cyclic symmetry groups on the sphere could also have similar constructions that would bear light on the case of S_{2n}; for example, if a nice construction was found for a diagram with polygyros symmetry group C_{nh}, the group S_{2n} is a subgroup of the polygyros group and so the construction could apply as well.

Chapter 8

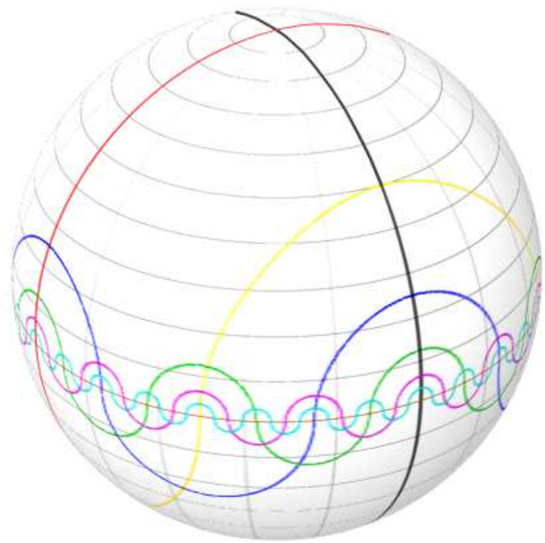
Other Symmetric Diagrams on the Sphere

This chapter contains discussions of a variety of different diagrams and constructions on the sphere that have interesting symmetry groups; they are collected here as they do not fit into the setting of other chapters. Starting with an examination of the spherical symmetries of one of the most well-known constructions for Venn diagrams, we then discuss a variety of related diagrams, and finish with a discussion of monochrome symmetry that includes constructions of diagrams on each of the platonic solids.

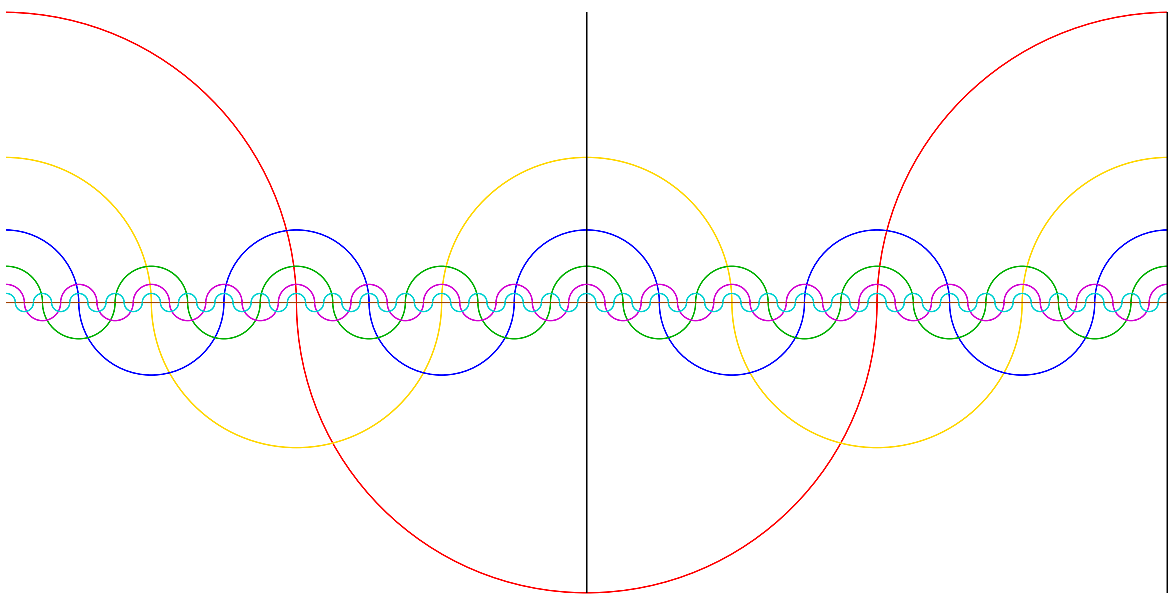
8.1 Edwards' Construction on the Sphere

In Section 3.2, Edwards' construction for an n-Venn diagram, for any n, was introduced [34, 37, 38], which is isomorphic to Grünbaum's construction from 1975 [56]—for simplicity we follow common usage and refer to the construction as *Edwards diagrams*. Edwards noted the possibility of observing interesting symmetries of the diagrams produced by embedding them on the sphere, and in this section we expand and formalize this idea.

Figure 8.1 shows the construction up to $n = 8$ curves on the sphere, and the cylindrical projection, drawn using circular arcs. The first three curves in the construction are the equatorial curve and the two orthogonal great circles passing through the poles, with the third curve considered to be the red curve. Each subsequent curve



(a) Diagram on the sphere



(b) Cylindrical projection, using semicircular arcs to construct curves for $n \geq 3$

Figure 8.1: The Edwards construction for an n-Venn diagram, drawn on the sphere

is a smaller and repeated copy of the previous, bisecting each region at a point on or close to the equator. Note that Figure 8.1(b) shows a planar projection which is not exactly a cylindrical projection as the second curve that passes through the poles (drawn in red) is drawn with circular arcs; this is to emphasize the "fractal" symmetry, taking advantage of the similarity of structure of the successive curves, disussed in Section 8.1.1.

On the sphere the diagram has reflective total symmetries across the orthogonal vertical planes containing the two curves passing through the poles. Thus, for the total symmetry group of the diagram we have $S_T = \mathsf{C}_{2v}$, which is isomorphic to the Klein four-group, the smallest non-cyclic group.

If the yellow curve is ignored in the construction there is a curve-preserving rotational symmetry (which is a total symmetry on all but the red and black curves) of order four (symmetry operation $R_{\vec{v},\pi/4}$) about the vertical axis through the poles. Note however that if the yellow curve is ignored the construction does not produce a Venn diagram for $n > 3$. The construction up to $n = 4$ curves, consisting of the three orthogonal great circles and the yellow curve in Figure 8.1, has a curve-preserving rotary reflection symmetry $G_{\vec{v},\pi/4}$, also about the polar axis of rotation.

On a more frivolous note, Edwards suggested drawing the construction on a tennis ball since the seam on a standard tennis ball is a nice sinusoidal curve that can be used as the fourth curve (yellow in Figure 8.1). Figure 8.2 shows the construction to $n = 6$ curves on a tennis ball.

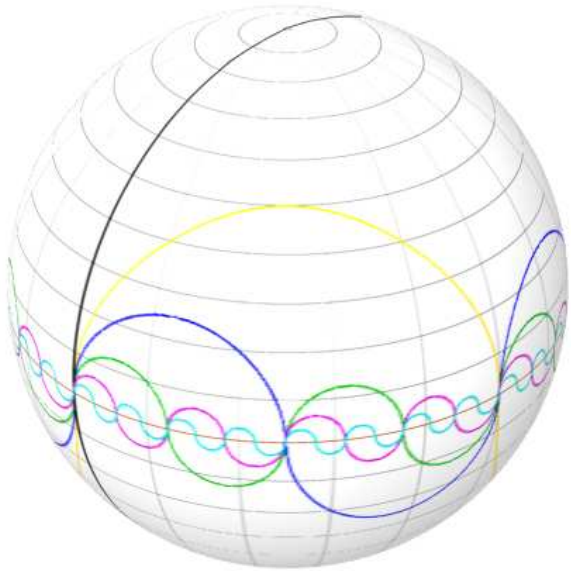
The Edwards construction also has an interesting variant, called *binary-form Edwards diagrams*. The name arises from their connections to Gray codes, which can be found in the diagram by considering the labelled dual graph—see Edwards' publications [32] and [34, Chapter 5] for more details. To create the diagrams, from the diagram in Figure 8.1, the ith curve is rotated about the polar axis by $\pi/2^{i-2}$ radians (in the cylindrical projection they are shifted horizontally, where the equatorial line

Figure 8.2: Edwards construction from Figure 8.1 drawn on a tennis ball

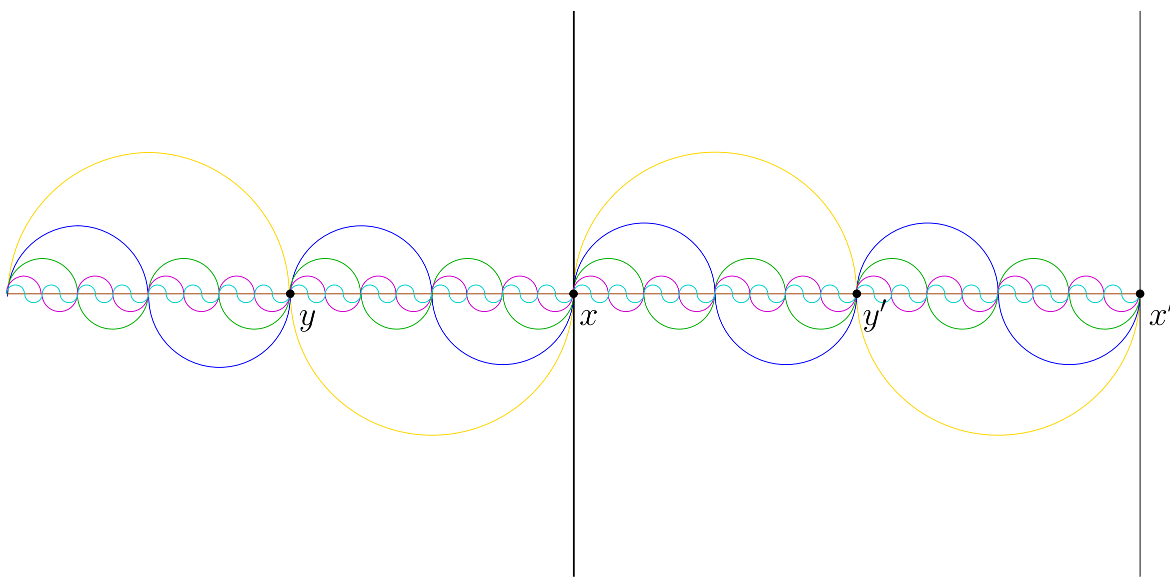
has length 2π). The third curve, coloured red in Figure 8.1, becomes identified with the second (black) curve, but the other curves remain distinct. The resulting diagram is shown in Figure 8.3.

On the sphere the binary-form diagram has three rotational total symmetries $R_{\vec{v},\pi}$ about orthogonal axes: the vertical (polar axis) coinciding with the z axis, and the x and y axes shown in Figure 8.3(b) passing through (x, x') and (y, y') respectively. These three orthogonal symmetries form a group of order four, as one rotational isometry is a composition of the other two, and they thus give total symmetry group $S_T = \mathsf{D}_2$ of order four.

Similar to Figure 8.2, Figure 8.4 shows the binary-form diagram drawn on a tennis ball, again with the yellow curve drawn as the tennis ball seam.



(a) Diagram on the sphere



(b) Cylindrical projection

Figure 8.3: Edwards construction modified to binary-form, drawn on the sphere

Figure 8.4: Binary-form Edwards construction from Figure 8.3 drawn on a tennis ball

8.1.1 Fractal Symmetries

A *fractal*, as (somewhat loosely) defined by Mandelbrot (the originator of the famous Mandelbrot Set fractal), is

> ...a rough or fragmented geometric shape that can be subdivided into parts, each of which is (at least approximately) a reduced-size copy of the whole. [88]

The concept has existed in mathematics for several hundred years in referring to self-similarity in geometric (and other) objects, such as the Sierpinski Triangle and the space-filling curve, and the term has now entered common usage in the sense Mandelbrot described.

In [34, Chapter 5], Edwards briefly discusses constructing the Edwards diagrams from trigonometric curves, and taking the construction to arbitrary, “fractal”, limits. The n-curve diagram in Figure 8.1(b) can be constructed from on the family of cosine curves $y = \cos(2^{i-2}x)/2^{i-2}$, $0 \leq x \leq 2\pi$, for $2 \leq i \leq n$. Similarly, the diagram in Figure 8.3(b) is given by the sine curves $y = \sin(2^{i-2}x)/2^{i-2}$, $0 \leq x \leq 2\pi$ for $2 \leq i \leq n$. In both cases curve C_1 is the only exception to the formulae in that it is given by the two vertical lines at $x = \pi$ and $x = 2\pi$.

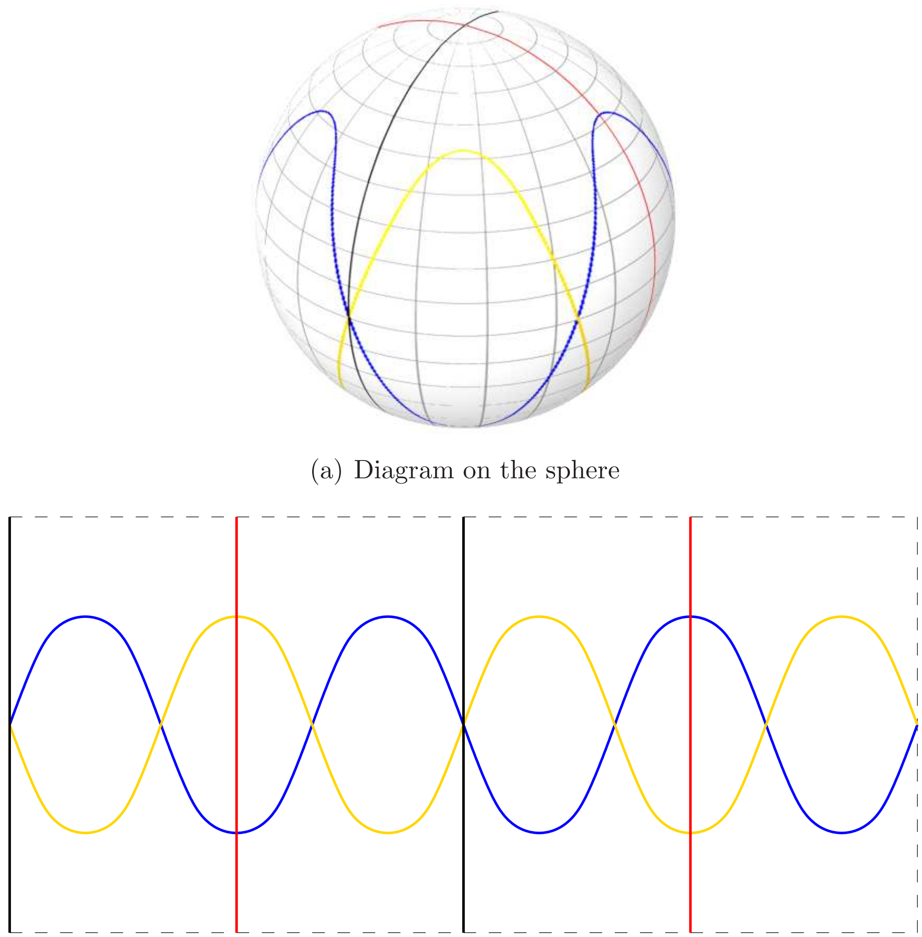
Since each curve C_i has the same equation once given the parameter i, these diagrams can be interpreted as having a fractal symmetry, in the following sense.

Consider continuing the diagram construction for Figure 8.1(b) to an infinite number of curves, to give the diagram $D_\infty = \{C_1, C_2, \ldots\}$ with C_i given by the formula described above. Then we have a new symmetry operation, distinct from those functions in Section 4.2, generating this fractal. Using the cylindrical projection (Figure 8.1(b)), consider the function $f(x, y) = \{(x/2, y/2)\ , (x/2 + \pi, y/2)\}$, with an appropriate exception for points on C_1, which is a curve-preserving symmetry operation taking curve C_i to curve C_{i+1} for $i \geq 2$. This operation shrinks each curve and copies it onto the next highest indexed. Given this symmetry we can consider D_∞ as being a fractal.

The mathematics of fractals are too complex to explain here (see [88], among others, for more details), and so we restrict our use of the concept to merely noting this fractal property of the diagram as shown. There are various ways of measuring fractals, the most common being the Hausdorff dimension and the box dimension; in this situation the fractal dimension of D_∞ is 1, according to either dimension [97].

8.2 A Different Construction with Antipodal Symmetry

In this section we introduce a construction, for any n, of an n-Venn diagram with some interesting symmetries on the sphere. Like some of the the diagrams in Chapter 5, these diagrams have antipodal symmetries, as well as some curve-preserving symmetries when $n \leq 4$.



(a) Diagram on the sphere

(b) Cylindrical projection of the diagram

Figure 8.5: A 4-Venn diagram with some fixed points in its curve-preserving symmetries

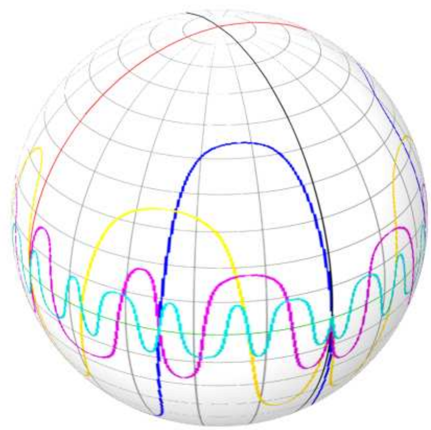
Figure 8.5 shows a 4-Venn diagram with an interesting aspect of its curve-preserving symmetry group. First, its total symmetries are a reflections across the plane containing the red curve passing through the poles, as well as symmetry under an antipodal map, giving the total symmetry group $S_T = \mathsf{C}_{2h}$. Adding in curve-preserving symmetry, the equatorial reflection gives the curve permutation (red black)(blue)(yellow), interesting in that it has two fixed points, which is different from most diagrams whose curve-preserving symmetry groups have either tended to be larger cycles or other groups with orbits of size at least two. The curve-preserving symmetry group is thus $S_C = \mathsf{D}_{2h}$ of order eight. Due to the asymmetry in how the red and black curves cross the blue and yellow curves it is clear that no curve-preserving symmetry will change the red and black curves.

This diagram is related to diagrams from another method of constructing an n-Venn diagram on the sphere for any n, which is nonisomorphic to the Edwards' construction (and other constructions isomorphic to that). After $n = 2$ each subsequent curve is arranged along the equatorial curve, dividing each existing region in two. The construction up to $n = 7$ is shown in Figure 8.6; it should be fairly clear how subsequent curves are added, with each smaller curve a smaller repeated copy of the previous.

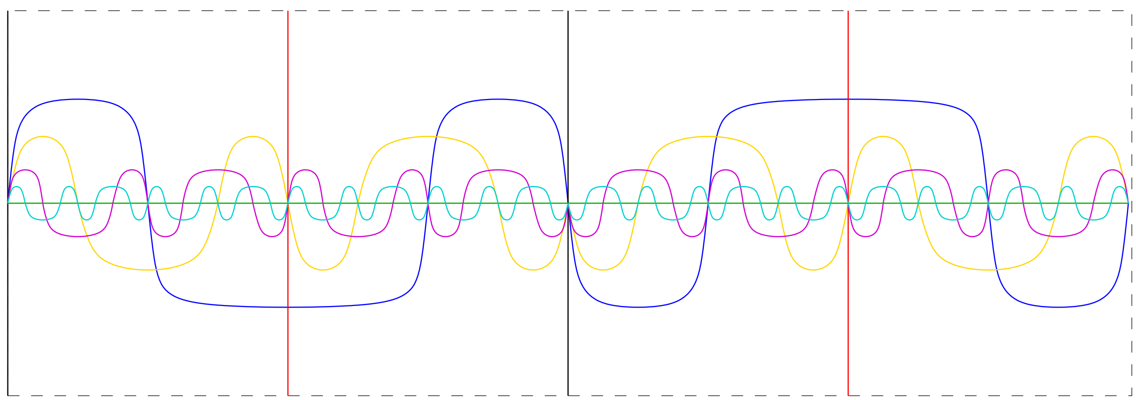
Any diagram from this construction for large n will have symmetry under the antipodal map, giving it the polydromic total symmetry group $S_T = \mathsf{S}_2$.

Unlike the polar symmetric and antipodally symmetric diagrams in Chapter 5, the dual graph of these diagrams is not easily constructed using chain decompositions; when each new curve is added each node in the dual graph is split into two nodes (according to a fairly simple rule) to generate the new dual.

In contrast to Figure 8.5, the diagram in Figure 8.7 is at first glance very similar but has a different symmetry. The Figure shows a 4-Venn diagram with the prismatic curve-preserving symmetry group $S_C = \mathsf{D}_{4h}$, of order 16. A reflection across the equatorial plane gives the curve permutation (yellow blue)(red)(black), whereas a rotation by $\pi/2$ around the polar axis gives the permutation (yellow blue)(red black), and there is a vertical reflection symmetry as well; the rotation symmetry gives an orbit of size four. However, unlike the diagram in Figure 8.5, this idea, if extended to a construction for an arbitrary number of curves, cannot create Venn diagrams if it is continued past four curves.



(a) Diagram on the sphere

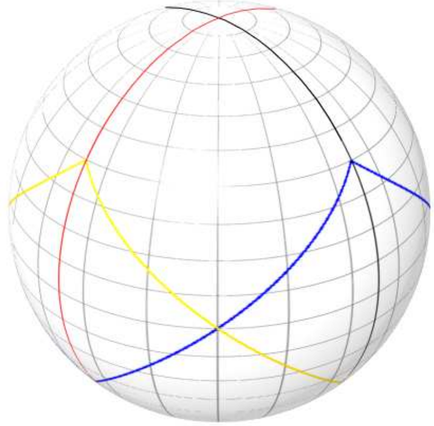


(b) Cylindrical projection of the diagram

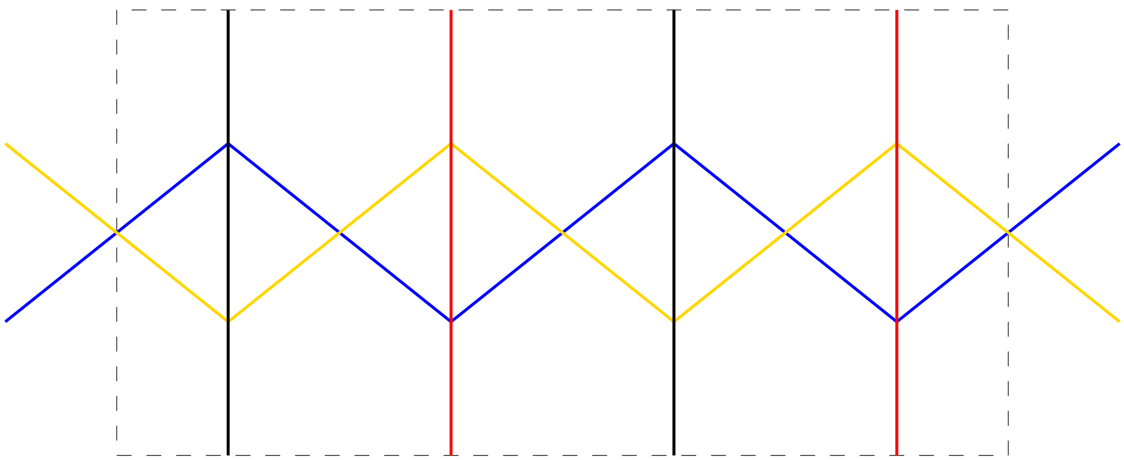
Figure 8.6: Construction for different kind of n-Venn diagram with antipodal symmetry

8.2.1 Related Three- and Four-curve Diagrams

We have already seen many different embeddings of the simple symmetric three-curve Venn diagram (Figure 3.6(a)) on the sphere and explored some of its symmetries; we will see more of this diagram in Section 8.3.2. The other symmetric three-curve diagram, in Figure 3.6(b), also has more symmetries when embedded on the sphere. Figure 8.8 shows this diagram embedded around the equator such that the arcs are symmetric about a reflection across the equator; there is also a monochrome symmetry by rotation $R_{2\pi/3}$ about a polar axis. This diagram then has monochrome symmetry group $S_M = \mathsf{D}_{3h}$, of order 12, while its curve-preserving symmetry group



(a) Diagram on the sphere



(b) Cylindrical projection of the diagram

Figure 8.7: 4-Venn diagram with curve-preserving symmetry group D_{4h}

is the dihedral group $S_C = \mathsf{D}_3$ of order six. Since the (planar) diagram is polar symmetric, we know from Section 4.5.6 that the spherical curve-preserving symmetry group is D_3 of order six.

A related three-curve diagram that is also similar to the four-curve diagrams of the previous section is shown in Figure 8.9; this diagram is a spherical embedding of one of the three-curve diagrams shown in Table 4.1 (Diagram # 3.4 from [106]). This diagram is interesting in that it has total symmetry group $S_T = \mathsf{D}_{1h}$ as there are two orthogonal axes of rotation, a polar axis and an equatorial axis through the vertices of degree six, that give total symmetries.

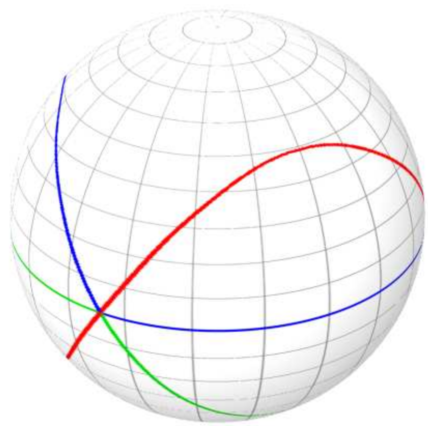
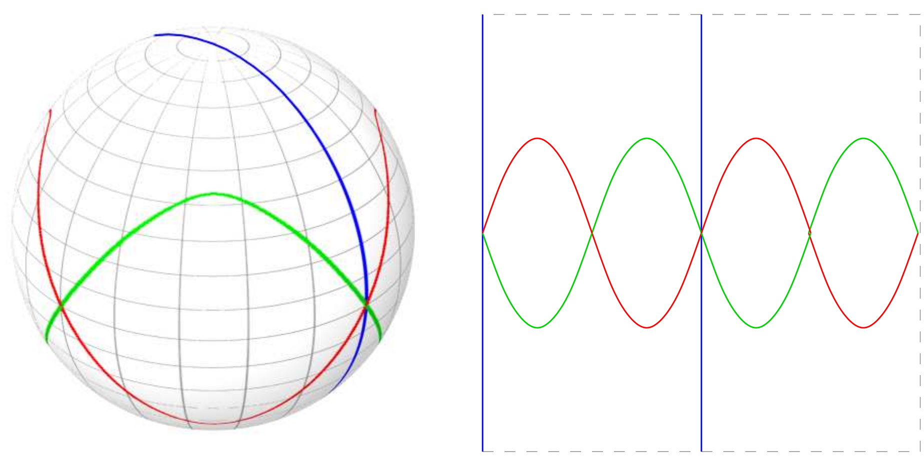


Figure 8.8: The non-simple symmetric three-curve Venn diagram on the sphere



(a) Diagram on the sphere (b) Cylindrical projection of the diagram

Figure 8.9: An embedding of a different three-curve diagram with prismatic symmetry

A three-curve diagram with the dihedral symmetry group under curve-preserving symmetries was shown in Figure 7.4, in the context of the CCR register being used to generate diagrams; the diagram also has reflective symmetries, giving it the symmetry group D_{3d} of order 12.

Similar to Figure 8.8 is the spherical embedding of the pseudo-symmetric four-curve Venn diagram of Figure 4.9, which we already observed has a relatively rich monochrome symmetry group (and no curve-preserving symmetries). When embedded on the sphere it has monochrome symmetry group $S_M = \mathsf{D}_{2h}$ of order eight, since

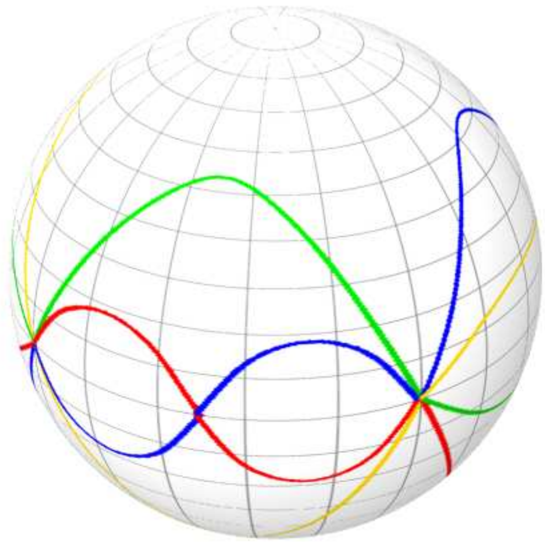


Figure 8.10: The pseudosymmetric four-curve Venn diagram on the sphere

there are three orthogonal planes of reflection; see Figure 8.10.

An even richer monochrome symmetry group is found by embedding some of the symmetric five- and seven-curve Venn diagrams on the sphere. In Table 4.1 we mentioned the five-curve diagram from [106] with monochrome symmetry group D_5 on the plane; on the sphere, as in Figure 8.11, it has monochrome symmetry group D_{5h} of order 20. The diagram in Figure 8.11 can be generated from the GKS construction (Section 3.3.2); the seven-curve diagram, referenced in Table 4.1, can be embedded in the same way to give a diagram on the sphere with monochrome symmetry group D_{7h} of order 28.

8.3 Monochrome Symmetries

In this section we will exhibit a selection of diagrams with interesting monochrome symmetry groups. In some respects monochrome symmetry is the least interesting of the three types of symmetries we have examined since it ignores curve colourings. However there are some interesting graphs that we have already seen with nontrivial monochrome symmetries, and the simple 3-Venn diagram, which we have already

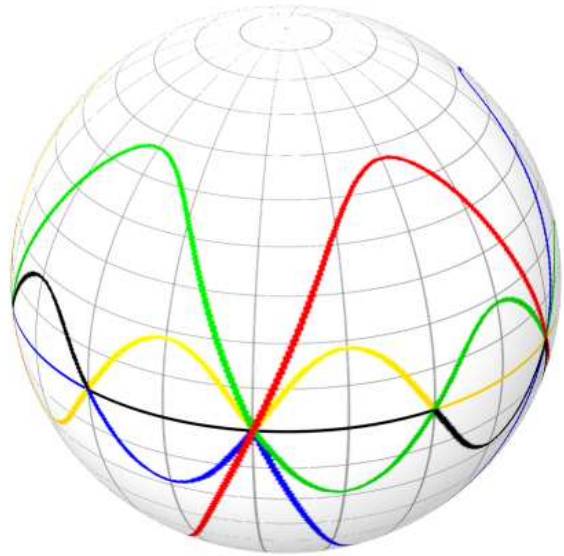


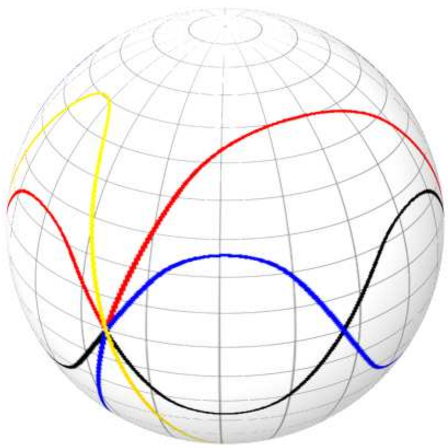
Figure 8.11: A symmetric five-curve Venn diagram on the sphere with rich monochrome symmetry group

seen is quite fruitful, leads to an interesting discussion of the Platonic solids when embedded on a tetrahedron.

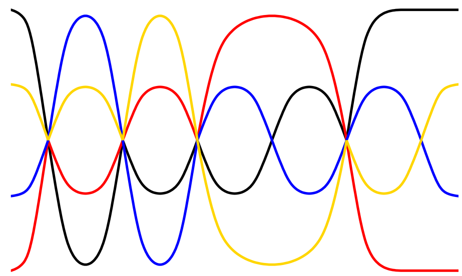
8.3.1 Diagrams from Chain Decompositions

As discussed in Section 3.3.1, the chains of the Greene Kleitman decomposition are symmetric, in the sense that every chain starts with an element of rank k and ends with an element of rank $n-k$, so it is balanced about the middle rank. This construction results in a symmetry if the diagram is embedded on the sphere.

If the chains are embedded vertically (on lines of longitude) on the sphere with all elements of middle rank placed at the same z coordinate $z = 0$ (the equator), so that all chains extend an equal distance above and below the equator, the resulting diagram will have reflective symmetry across the plane $z = 0$ since the structure of the dual diagram, and thus the diagram itself (ignoring curve colours), is reflected across this line. Thus, by applying the construction of [54] in Section 3.3.1 to create monotone n-Venn diagrams of any n, the resulting diagrams have a monochrome symmetry group $S_M = \mathsf{C}_{1h}$. This is apparent in Figure 8.12, which is an embedding on the sphere of the 4-curve Venn diagram from Figure 3.8. Note that this particular diagram for $n = 4$, because the construction has a small number of curves, happens to also have another reflective symmetry across a vertical plane through the sphere, giving a monochrome symmetry group for this diagram of $S_M = \mathsf{D}_{1h}$ of order four.



(a) Diagram on the sphere



(b) Diagram on the plane

Figure 8.12: Monochrome symmetry of monotone 4-Venn diagram from [54], by reflection across equator

This symmetry is not evident in the planar diagrams generated by the construction, because the two regions corresponding to the empty set and complete set (labelled 0^n and 1^n in Figure 3.8) do not share exactly complementary roles: whereas on the sphere the empty and complete set both form closed regions on opposite sides of the sphere, on the plane the complete set is contained within curves separating it from the unbounded outer set. In Figure 3.8 this is apparent in that the symmetry described above pertains to all of the structure of the diagram except the edges that "wrap around" the ends of the diagram.

This monochrome symmetry is not present in the alternate constructions from Chapter 5, since the chains in those decompositions are not symmetric about the middle rank.

8.3.2 Diagrams on the Platonic Solids

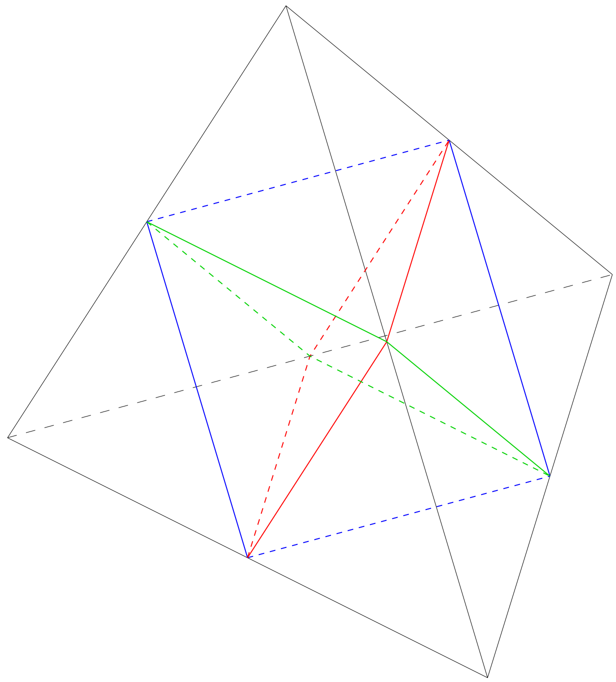
In this section we exhibit some diagrams that realize some of the symmetry groups in Table 2.2 that we have not seen otherwise, notably the groups with multiple axes of rotation which are realized by the Platonic solids. The five Platonic solids, also called regular solids or regular polyhedra, are the five convex three-dimensional polyhedra that each have all internal angles equal and all vertices are equal degree.

This section is joint work with Brett Stevens.

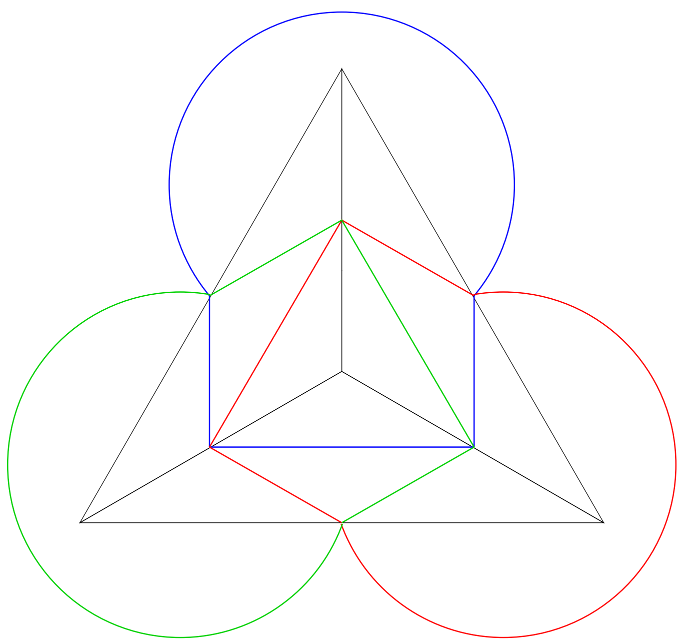
The regular polyhedra each have the property that an enclosing sphere can be constructed with all vertices of the polyhedron lying on the sphere. We can thus draw a diagram with the polyhedron's symmetry group on the surface of the polyhedron, to better exhibit the symmetries involved, and the diagram can be projected from the polyhedron onto the surface of the containing sphere, preserving all symmetries, as follows: for each point on the diagram, project it onto the surface of the containing sphere along a ray through that point from the centre of the sphere, as if a point light source at the centre of the sphere were casting a shadow of the diagram onto the surface of the sphere.

The simple three-curve diagram that has been so useful throughout this work can also be exploited to realize symmetry groups of some Platonic solids, as follows. By embedding vertices of the diagram on the midpoints of edges, the diagram maps onto the tetrahedron, with four faces containing the tetrahedron's vertices and four faces occupying the central quarter of each face. The diagram in Figure 8.13 has monochrome symmetry group $S_M = \mathsf{T}_d$, called achiral tetrahedral symmetry. By Corollary 1, since the diagram is simple, we have the fact that the curve-preserving symmetry group S_C is also T_d.

The tetrahedral symmetry group T_d, as well as the pyritohedral group T_h, are both subgroups of the octahedral group O_h, and in fact the same simple three-curve Venn diagram can also be embedded to realize the octahedral group, as in Figure 8.14. By



(a) Diagram on the tetrahedron



(b) Stereographic projection from a vertex

Figure 8.13: Spherical 3-Venn diagram showing tetrahedral symmetry group T_d

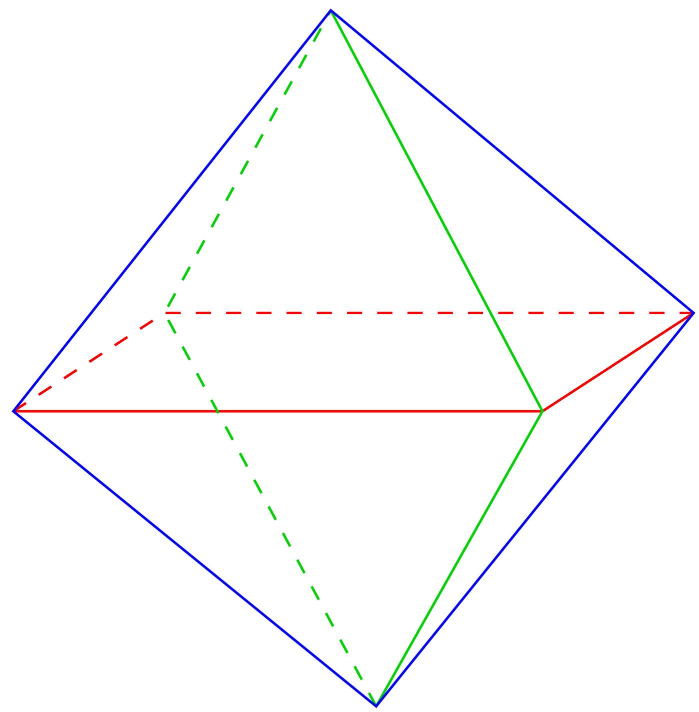


Figure 8.14: 3-Venn diagram showing octahedral curve-preserving symmetry group O_h, with curves forming edges of the octahedron

embedding edges as edges and vertices as vertices of the octahedron, this diagram is perhaps one of the more natural realizations of multiaxial symmetry groups. Again we have that since the diagram is simple, the monochrome symmetry group and curve-preserving symmetry group are both $S_M = S_C = \mathsf{O}_h$.

The octahedron and the cube are dual solids, meaning that if a cube was constructed with vertices in the centres of the faces of the octahedron, the resulting cube has the same symmetries. Thus, as we saw in Section 6.5, if the 3-Venn diagram as in Figure 8.14 was projected onto such a cube it would also realize the same symmetries; see Figure 6.5 on page 153.

For the largest symmetry group in Table 2.2, the icosahedral group, the three-curve diagram will no longer suffice. On the surface of the dodecahedron, which has 12 faces and degree-three vertices, a five-curve Venn diagram is required, since four curves provide 16 faces, and there is no combination of features on the dodecahedron which sum to 16. A 5-Venn diagram has 32 faces, which map naturally to the 12 faces and 20 vertices of the dodecahedron, suggesting an approach similar to the diagram in Figure 8.13, in which some faces enclose vertices and some remain on the faces, with vertices embedded on the midpoints of edges. All vertices must be the same degree, to preserve symmetry. With all vertices of degree four, the 12 faces of the dodecahedron give 60 edges in the diagram. Since all vertices are degree four, the resulting diagram must be simple. Without loss of generality pick any 5-face and label it as the empty region, and the five regions surrounding it must then be enclosed by the five different curves. Given this each curve can be completely labelled and the entire diagram built, but it is easy to check that it does not form a Venn diagram as some regions are missing and an extra curve is necessary to fix the vertex counts. See Figure 8.15 for a stereographic projection of this diagram—the empty set is the external face, which then forces the five neighbouring edges to be the five different curves, which forces the colours of all other edges belonging to those curves, and the extra curve appears, shown in purple.

Thus a simple 5-Venn diagram with monochrome icosahedral symmetry (and thus curve-preserving also) cannot be drawn. However, a solution presents itself with some relaxation of the types of diagram we allow. By allowing several curves to coexist at the same point, a non-simple diagram can be drawn, with three or more curves passing through a vertex that only has degree four. Such diagrams relax the condition from Section 2.1 that curves intersect at a finite number of points.

Definition. A diagram is *infinitely-intersecting* if curves meet in infinitely many points.

Infinitely-intersecting diagrams lack some of the nice graph-theoretic properties of other diagrams, as a given edge in the diagram can contain multiple curves, and thus labels of the two adjacent regions on either side of the edge may differ in more than one bit position. However they can be a useful gain in flexibility; for other work

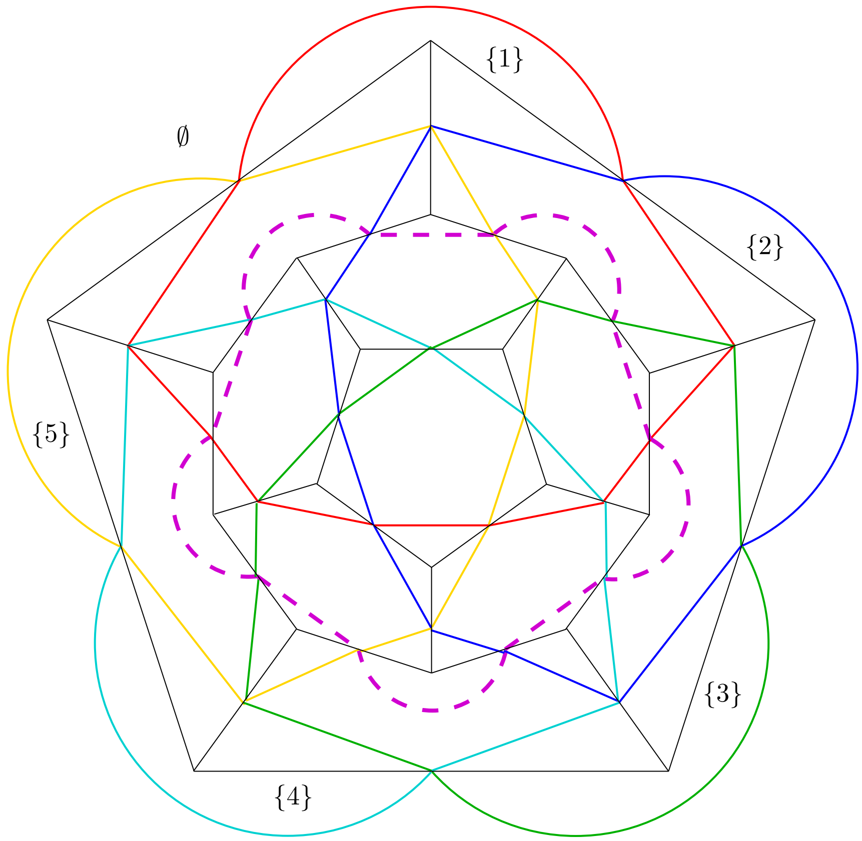


Figure 8.15: An attempt at drawing a spherical simple 5-Venn diagram on the dodecahedron with monochrome icosahedral symmetry group I_h fails; an extra curve (shown in purple) appears

using them, a construction using de Bruijn cycles and infinitely-intersecting diagrams is in [52, pp. 17, 20], a construction using polyominoes to create Venn diagrams is [20], and some other applications are in [106, "Generalizing Venn Diagrams"].

Figure 8.16 shows a 5-Venn diagram with monochrome symmetry group $S_M = \mathsf{I}_h$ on the sphere. All vertices are degree-four for the purposes of the monochrome symmetry, but from the diagrammatic perspective three curves pass through each vertex.

This diagram also has curve-preserving rotational isometries, as is more evident in Figure 8.17, showing the stereographic projection (compare with Figure 8.15), and polar symmetry (see Section 4.5.6), giving it the curve-preserving symmetry group

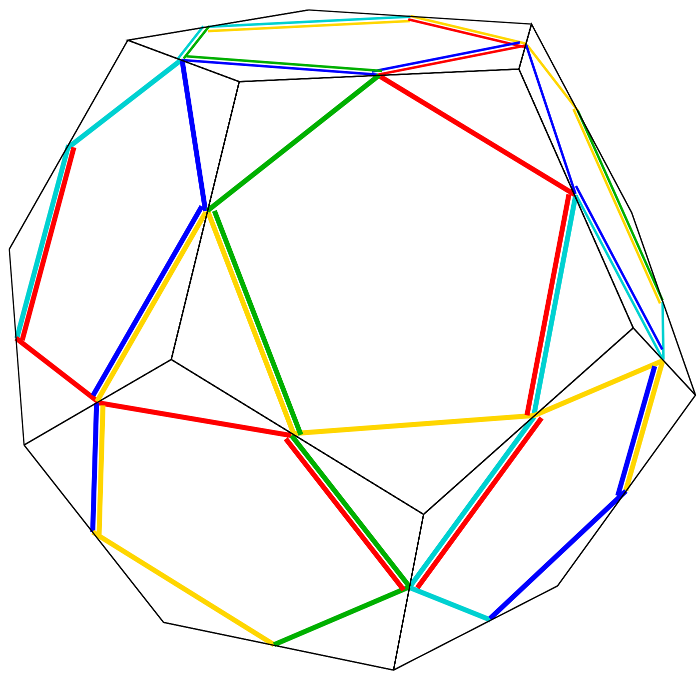


Figure 8.16: Spherical infinitely-intersecting 5-Venn diagram drawn on the dodecahedron with monochrome symmetry group I_h; parallel multiedges are colocated

$S_C = \mathsf{D}_5$.

The icosahedron and the dodecahedron are dual solids, and so if the 5-Venn diagram in Figure 8.16 was projected onto an icosahedron it would also realize the same symmetries. In such a diagram, each 5-face of the diagram would surround a vertex of the icosahedron and each 3-face would embed in the centre of a face, similarly to Figure 8.13.

These diagrams all realize the achiral, or nonoriented, groups of the Platonic solids, but it is easy to see that by adding widgets to the specific curves (for example, small widgets shaped like "⏌⟍_" added into the midpoint of each edge all oriented with the same "handedness") the chiral versions of these groups can also be realized. Thus we also can realize the chiral tetrahedral group T of order 12, chiral octahedral group O of order 24, and chiral icosahedral group I of order 60.

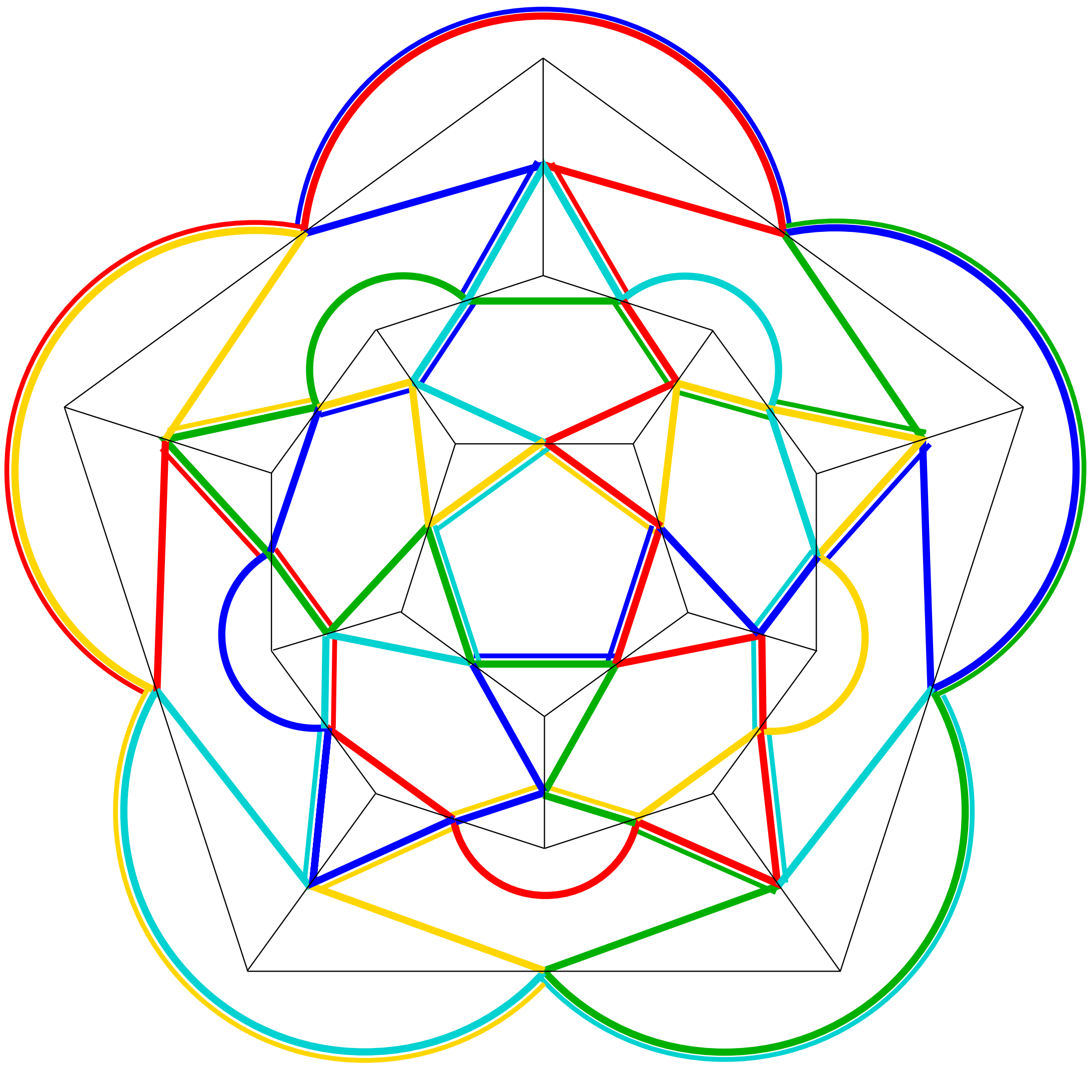


Figure 8.17: Stereographic projection of infinitely-intersecting 5-Venn diagram in Figure 8.16

Open Questions

It is interesting to speculate on the possibilities for creating diagrams with symmetry groups of Platonic solids in higher dimensions. There are six regular polyhedra in four dimensions, three in five dimensions, and three in all higher dimensions. Our work in Chapter 6 created diagrams with, as in Figure 8.14, the total symmetries of an octahedron and subgroups thereof, so we have already realized the total symmetry group of one of the regular polyhedra for higher dimensions.

Chapter 9

Conclusions

Our research into building and embedding diagrams, both Venn and otherwise, on the sphere has provided many useful results and interesting diagrams. Table 9.1 encapsulates the contributions of this thesis by displaying what is currently known about the symmetry groups realizable on the sphere by Venn diagrams.

Table 9.1 is an extended version of Table 2.2. The first column gives the standard name, as used in Table 2.2, of the class of symmetry, and for the prismatic classes, with parameter n, each class shows different entries, corresponding to different values of n. For each symmetry group, the standard group notation and its order is shown, and then we show a different entry for each of the three types of colour symmetry discussed in Chapter 4. Each entry in the table gives figure references to where the various diagrams are displayed elsewhere in this work, and section references if they are discussed (without a figure) in the context of other work, and we also include the more significant conjectures. Recall Lemma 7 in Section 4.3.1, which tells us that, due to the containment properties of the total, curve-preserving and monochrome symmetry groups, any figure satisfying a total symmetry group also satisfies the corresponding curve-preserving and monochrome groups, and similarly curve-preserving symmetries satisfy monochrome groups.

The table shows that we have succesfully found at least one diagram in each of the 14 classes of spherical symmetry, many with several different groups, and many with different types of colour symmetry. All of the constructions and diagrams are new to this work, except for the following diagrams, which are all used in new ways and new embeddings:

- GKS and related constructions realizing the groups $\mathsf{C_n}$, C_{1h}, C_{nh}
- pseudosymmetric diagrams realizing the groups C_{2h}, C_{3h}
- Edwards constructions realizing the groups C_{2v}, D_2
- polar symmetric diagrams realizing the groups D_5, D_7
- three-curve diagrams realizing the groups D_3, D_{1h}
- various uses of the simple 3-curve diagram, such as realizing the groups T_d and O_h

This work has all been based on the framework begun in Chapter 4. Our work has also introduced some new ways of constructing diagrams that we hope can be further explored to find more and larger diagrams with rich symmetry groups on the sphere or other surfaces.

9.1 Open Questions

While we feel that the situation on the plane with regards to symmetry is quite well explored, on the sphere there are many large areas in which we have only begun to investigate what is possible; even restricting ourselves to Venn diagrams, as opposed to examining independent families or Euler diagrams more generally, we have found many diagrams in just this examination of the terrain.

We hope that this thesis has provided a foundation for further research on chain decompositions and symmetry groups of diagrams, Venn and otherwise, on the sphere

Table 9.1: Known Venn diagrams realizing the finite groups on the sphere

Name	**Group Notation**	**Order**	**Total Symmetry**	**Curve-preserving Symmetry**	**Monochrome Symmetry**
polytropic	C_1	2	$n > 1$: RSCD construction (Sec 5.5)		
(rotational)	C_n	n		n prime, ≥ 2: GKS construction (Sec 3.3.2)	
polygyros	C_{1h}	2			$n > 1$: GKS monotone construction (Sec 3.3.1)
	C_{2h}	4	$n = 4$: Fig 8.5		$n = 4, 8$: pseudosymmetric diagrams (Sec 4.3, [107])
	C_{3h}	6			$n = 9$: pseudosymmetric diagram (Sec 4.3, [107])
	C_{nh}	$2n$		add widgets to any diagram with D_{nh}	n prime, ≥ 3: GKS construction (Sec 3.3.2)
polyscopic	C_{2v}	2	$n \geq 1$: Edwards construction (Fig 8.1) $n = 4$: Fig 8.7		

Table 9.1: (continued)

Name	**Group Notation**	**Order**	**Total Symmetry**	**Curve-preserving Symmetry**	**Monochrome Symmetry**
(pyramidal)	C_{nv}	$2n$		add widgets to any diagram with D_{nh}	
polyditropic	D_2	8	$n \geq 1$: Edwards binary-form (Fig 8.3)		
(dihedral)	D_3	6		$n = 3$: Fig 8.8	
	D_5	10		$n = 5$: simple, Fig 4.15	
	D_7	14		$n = 7$: simple, Sec 4.5.6	
polydiscopic	D_{1h}	4	$n = 3$: Fig 8.9		$n = 4$: Fig 8.12
(prismatic)	D_{2h}	8	$n = 2$: Fig 7.23	$n = 4$: Fig 8.5	$n = 4$: Fig 8.10
	D_{3h}	12			$n = 3$: Fig 8.8
	D_{4h}	16		$n = 4$: Fig 8.7	
	D_{5h}	20			$n = 5$: Fig 8.11
	D_{7h}	28			$n = 7$: Sec 8.2.1
polydigyros	$\mathsf{D}_{1d} = \mathsf{C}_{2h}$	4	see C_{2h}		

Table 9.1: (continued)

Name	**Group Notation**	**Order**	**Total Symmetry**	**Curve-preserving Symmetry**	**Monochrome Symmetry**
(anti-prismatic dihedral)	D_{3d}	12		$n = 3$: Fig 7.4 (Sec 8.2.1)	
polydromic	S_2	2	$n > 1$: ASCD construction (Sec 5.5) $n > 1$: antipodal construction (Fig 8.6)		
	S_8	8		$n = 4$: CCR example (Fig 8.6)	
	S_{10}	10		$n = 5$: Fig 7.16	
	S_{16}	16		$n = 8$: CCR constructions (Sec 7.5.3)	
	S_{2n}	$2n$		$n = 2^k$, prime, or Poulet number: conjectured (Sec 7.5.3)	

Table 9.1: (continued)

Name	**Group Notation**	**Order**	**Total Symmetry**	**Curve-preserving Symmetry**	**Monochrome Symmetry**
chiral tetrahedral	T	12		add widgets to diagram with T_d (Sec 8.3.2)	
tetrahedral	T_d	24		$n = 3$: Fig 8.13	
pyritohedral	T_h	24		add widgets to diagram with O_h (Sec 8.3.2)	
chiral octahedral	O	24		add widgets to diagram with O_h (Sec 8.3.2)	
octahedral	O_h	48		$n = 3$: Diagram Δ_3 (Sec 6.5), $n = 3$: Fig 8.14	
chiral icosahedral	I	60			add widgets to diagram wih I_h (Sec 8.3.2)
icosahedral	I_h	120			$n = 5$: Fig 8.16 (infinitely intersecting)

and other higher-dimensional surfaces. We have mentioned some more context-specific open questions in the final sections of each chapter. In Table 4.1 the current knowledge was displayed on what symmetry groups on the plane are achievable with Venn diagrams with different types of colour symmetry. This research began with the question: what are all of the symmetry groups that are realizable on the plane? Thus inspired, we have progressed to examining other types of symmetry, symmetries on the sphere, and finally to higher-dimensional diagrams. Referring again to Table 9.1, it is interesting to ask what other chain decompositions, diagram constructions, and larger diagrams can be found with these symmetry groups. Furthermore we can ask what other constructions can be envisioned for chain decompositions and diagrams, Venn or otherwise, on other surfaces other than the plane, such as the projective plane or the torus, and investigate what symmetries and colour symmetries such diagrams would have.

Bibliography

[1] Martin Aigner. Lexicographic matching in Boolean algebras. *J. Combinatorial Theory*, 14:187–194, 1973.

[2] Daniel E. Anderson and Frank L. Cleaver. Venn-type diagrams for arguments of n terms. *Journal of Symbolic Logic*, 30:113–118, 1965.

[3] M. A. Armstrong. *Groups and Symmetry*. Springer-Verlag, 1988.

[4] D. Barker-Plummer and S. Bailin. The role of diagrams in mathematical proofs. *Machine GRAPHICS and VISION, (Special Issue on Diagrammatic Representation and Reasoning)*, 6(1):25–56, 1997.

[5] Margaret E. Baron. A note on the historical development of logic diagrams: Leibniz, Euler, and Venn. *Mathematical Gazette*, 53:113–125, 1969.

[6] Florence Benoy and Peter Rodgers. A study into the comprehension of Euler diagrams. Technical Report 14-04, University of Kent, Computing Laboratory, July 2004.

[7] Edmund C. Berkeley. Boolean algebra and applications to insurance. *The Record of the American Institute of Actuaries*, 26:373–414, 1937.

[8] Brian H. Bowditch. *A Course on Geometric Group Theory*. MSJ Memoirs. Mathematical Society of Japan, 2006.

[9] L. J. Bowles. Logic diagrams for up to n classes. *Mathematical Gazette*, 55:370–373, 1971.

[10] F. J. Budden. *The Fascination of Groups*. Cambridge University Press, 1972.

[11] B. Bultena, B. Grünbaum, and F. Ruskey. Convex drawings of intersecting families of simple closed curves. In *11th Canadian Conference on Computational Geometry*, pages 18–21, 1999.

[12] R. P. Burn. *Groups: A Path to Geometry*. Cambridge University Press, 1985.

[13] H. Burzlaff and H. Zimmermann. Point-group symbols. In Hahn [61], chapter 12.1, pages 818 – 820.

[14] Tao Cao. Computing all the simple symmetric monotone Venn diagrams on seven curves. Master's thesis, University of Victoria, 2001.

[15] Kevin Cattell, Frank Ruskey, Joe Sawada, Micaela Serra, and C. Robert Miers. Fast algorithms to generate necklaces, unlabeled necklaces, and irreducible polynomials over GF(2). *J. Algorithms*, 37(2):267–282, 2000.

[16] Kiran B. Chilakamarri, Peter Hamburger, and Raymond E. Pippert. Hamilton cycles in planar graphs and Venn diagrams. *J. Combin. Theory Ser. B*, 67(2):296–303, 1996.

[17] Kiran B. Chilakamarri, Peter Hamburger, and Raymond E. Pippert. Venn diagrams and planar graphs. *Geometriae Dedicata*, 62:73–91, 1996.

[18] Kiran B. Chilakamarri, Peter Hamburger, and Raymond E. Pippert. Analysis of Venn diagrams using cycles in graphs. *Geometriae Dedicata*, 82:193–223, 2000.

[19] Stirling Chow and Frank Ruskey. Searching for symmetric Venn diagrams. In *Workshop on Computational Graph Theory and Combinatorics*, pages 37–38, Victoria, 1999. Extended abstract.

[20] Stirling Chow and Frank Ruskey. Minimum area Venn diagrams whose curves are polyominoes. *Math. Mag.*, 80(2):91–103, 2007.

[21] Barry Cipra, Peter Hamburger, and Edit Hepp. Aesthetic aspects of Venn diagrams. In *Proceedings of the 2005 Bridges Conference on Mathematical Connections in Art, Music and Science (July 31-August 3, 2005)*, pages 339–342, Banff, Canada, 2005.

[22] S. N. Collings. Further logic diagrams in various dimensions. *The Mathematical Gazette*, 56(398):309–310, December 1972.

[23] L. Couturat, editor. *La Logique de Liebniz*. Paris, 1901. In French.

[24] L. Couturat, editor. *Opuscules et Fragments Inèdits de Liebniz*. Paris, 1901. In French.

[25] H. S. M. Coxeter. *Introduction to Geometry*. Wiley, New York, second edition, 1969.

[26] H. S. M. Coxeter. Regular and semi-regular polyhedra. In M. Senechal and G. Flecks, editors, *Shaping Space: a polyhedral approach*. Birkhauser-Boston, New York, 1988.

[27] H. S. M. Coxeter and W. O. J. Moser. *Generators and Relations for Finite Groups*. Springer-Verlag, Berlin, second edition, 1965.

[28] N. G. de Bruijn, Ca. van Ebbenhorst Tengbergen, and D. Kruyswijk. On the set of divisors of a number. *Nieuw Arch. Wiskunde (2)*, 23:191–193, 1951.

[29] R. P. Dilworth. A decomposition theorem for partially ordered sets. *Ann. of Math. (2)*, 51:161–166, 1950.

[30] David S. Dummit and Richard M. Foote. *Abstract Algebra*. Wiley, third edition, 2004.

[31] Herbert Edelsbrunner, Leonidas Guibas, János Pach, Richard Pollack, Raimund Seidel, and Micha Sharir. Arrangements of curves in the plane—topology, combinatorics, and algorithms. *Theoretical Computer Science*, 92(2):319–336, 1992.

[32] A. W. F. Edwards. Rotatable Venn diagrams. *Mathematics Review*, (2):19–21, February 1992.

[33] A. W. F. Edwards. The sevenfold symmetric Venn diagrams. In *XVIIth International Biometric Conference Proceedings*, volume 2, page 238, Hamilton, Ontario, 1994.

[34] A. W. F. Edwards. *Cogwheels of the Mind: The Story of Venn Diagrams*. The John Hopkins University Press, 2004.

[35] Anthony Edwards. Seven-set Venn diagrams with rotational and polar symmetry. *Combinatorics, Probability, and Computing*, 7:149–152, 1998.

[36] Anthony Edwards. An eleventh-century Venn diagram. *BSHM Bulletin*, 21:119–121, 2006.

[37] Anthony W. F. Edwards. Venn diagrams for many sets. In *Bulletin of the International Statistical Institute, 47th Session*, pages 311–312, Paris, 1989. Contributed Papers, Book 1.

[38] Anthony W. F. Edwards. Venn diagrams for many sets. *New Scientist*, 7:51–56, January 1989.

[39] Leonard Euler. *Lettres ã une Princesse d'Allemagne sur Divers Sujets de Physique et de Philosophie*. Académie Imperiale des Sciences, St. Petersburg, 1768–1772. In French. Translated by Sir David Brewster, Edinburgh, W & C Tait, and Longman et al., Vol. 1., 1823.

[40] Herbert Federer. *Geometric Measure Theory*. Springer, 1996.

[41] J. C. Fisher, E. L. Koh, and B. Grünbaum. Diagrams Venn and how. *Mathematics Magazine*, 61(1):36–40, 1988.

[42] Jean Flower and John Howse. Generating Euler diagrams. In *Diagrammatic representation and inference (Callaway Gardens, GA, 2002)*, volume 2317 of *Lecture Notes in Comput. Sci.*, pages 61–75. Springer, Berlin, 2002.

[43] Jean Flower, John Howse, and John Taylor. Nesting in Euler diagrams: syntax, semantics and construction. *Software and System Modeling*, 3(1):55–67, 2004.

[44] H. Fredricksen and I. J. Kessler. An algorithm for generating necklaces of beads in two colors. *Discrete Mathematics*, 61:181–188, 1986.

[45] H. Fredricksen and J. Maiorana. Necklaces of beads in k colours and k-ary de Bruijn sequences. *Discrete Mathematics*, 23:207–210, 1978.

[46] Harold Fredricksen. A survey of full length nonlinear shift register cycle algorithms. *SIAM Rev.*, 24(2):195–221, 1982.

[47] Joseph A. Gallian. *Contemporary Abstract Algebra*. Houghton Mifflin, New York, 2006. 6th edition.

[48] Martin Gardner. *Logic Machines and Diagrams*. Harvester Press, Brighton, Sussex, second edition, 1958.

[49] Martin Gardner. Curious maps. In *Time Travel and Other Mathematical Bewilderments*, chapter 15. W. H. Freeman, 1988.

[50] Andrew Glassner. Venn and Now. *IEEE Computer Graphics and Applications*, 23(4):82–95, July/August 2003.

[51] Solomon W. Golomb. *Shift Register Sequences*. Holden-Day, San Francisco, 1967.

[52] R. Graham, D. Knuth, and O. Patashnik. *Concrete Mathematics: A Foundation for Computer Science*. Addison-Wesley, second edition, 1994.

[53] C. Greene and D. Kleitman. Strong versions of Sperner's theorem. *Journal of Combinatorial Theory*, 20(1):80–88, 1976.

[54] J. Griggs, C. E. Killian, and C. Savage. Venn diagrams and symmetric chain decompositions in the boolean lattice. *The Electronic Journal of Combinatorics*, 11(1), 2004. Article R2 (online).

[55] Ralph P. Grimaldi. *Discrete and Combinatorial Mathematics: An Applied Introduction*. Addison-Wesley, 1994. 3rd edition.

[56] B. Grünbaum. Venn diagrams and independent families of sets. *Mathematics Magazine*, 48:12–23, 1975.

[57] B. Grünbaum. On Venn diagrams and the counting of regions. *The College Mathematics Journal*, 15(5):433–435, 1984.

[58] B. Grünbaum. Venn diagrams I. *Geombinatorics*, 1:5–12, 1992.

[59] B. Grünbaum. Venn diagrams II. *Geombinatorics*, 2:25–31, 1992.

[60] B. Grünbaum. The search for symmetric Venn diagrams. *Geombinatorics*, 8:104–109, 1999.

[61] Th. Hahn, editor. *International Tables for Crystallography Volume A: Space-group symmetry*, chapter 12, pages 818 – 834. Springer Netherlands, fifth edition, 2002.

[62] P. Hamburger, Gy. Petruska, and A. Sali. Saturated chain partitions in ranked partially ordered sets, and non-monotone symmetric 11-Venn diagrams. *Studia Sci. Math. Hungar.*, 41(2):147–191, 2004.

[63] P. Hamburger and A. Sali. Symmetric 11-Venn diagrams with vertex sets $231, 242, \ldots, 352$. *Studia Sci. Math. Hungar.*, 40(1-2):121–143, 2003.

[64] Peter Hamburger. Doodles and doilies, non-simple symmetric Venn diagrams. *Discrete Mathematics*, 2-3(257):423–439, 2002. Kleitman and combinatorics: a celebration (Cambridge, MA, 1999).

[65] Peter Hamburger. Pretty drawings. More doodles and doilies, symmetric Venn diagrams. *Util. Math.*, 67:229–254, 2005.

[66] Peter Hamburger and Edit Hepp. Symmetric Venn diagrams in the plane: The art of assigning a binary bit string code to planar regions using curves. *Leonardo*, 38(2):125–132, April 2005.

[67] Peter Hamburger and Raymond E. Pippert. Simple, reducible Venn diagrams on five curves and Hamiltonian cycles. *Geometriae Dedicata*, 68(3):245–262, 1997.

[68] Peter Hamburger and Raymond E. Pippert. Venn said it couldn't be done. *Mathematics Magazine*, 73(2):105–110, 2000.

[69] Peter Hamburger and Attila Sali. 11-doilies with vertex sets of sizes $275, 286, \ldots, 462$. *AKCE Int. J. Graphs Comb.*, 1(2):109–133, 2004.

[70] Georges Hansel. Sur le nombre des fonctions booléennes monotones de n variables. *C. R. Acad. Sci. Paris Sér. A-B*, 262:A1088–A1090, 1966. In French.

[71] G. H. Hardy and E. M. Wright. *Introduction to the Theory of Numbers*. Clarendon Press, Oxford, 4th edition, 1960.

[72] D. Henderson. Venn diagrams for more than four classes. *American Mathematical Monthly*, 70(4):424–426, 1963.

[73] Michael Henle. *A Combinatorial Introduction to Topology*. Dover, 1979.

[74] Harold Hilton. *Mathematical Crystallography and the Theory of Groups and Movements*. Dover, New York, 1963.

[75] C. Iwamoto and G. T. Toussaint. Finding Hamiltonian circuits in arrangements of Jordan curves is NP-complete. *Information Processing Letters*, 52(4):183 – 189, 1994.

[76] M. Jamnik, A. Bundy, and I. Green. On automating diagrammatic proofs of arithmetic arguments. *Journal of Logic, Language, and Information*, 8(3):297–321, 1999.

[77] M. A. Jaswon and M. A. Rose. *Crystal Symmetry: Theory of Colour Crystallography*. Ellis Horwood, Chichester, 1983.

[78] H. Kahane and P. Tidman. *Logic and Philosophy: A Modern Introduction*. Wadsworth, Belmont, 1995.

[79] Andrew Kepert. Illustration of a typical member of each of 7 infinite families of 3D point groups. `http://en.wikipedia.org/wiki/Image:Uniaxial.png`, 2006. [Online; accessed 28-August-2007].

[80] Felix Klein. *Lectures on the Icosahedron and the Solution of Equations of the Fifth Degree*. Dover Publications Inc., New York, N.Y., revised edition, 1956. Translated by George Gavin Morrice.

[81] Israel Kleiner. The evolution of group theory: a brief survey. *Mathematics Magazine*, 59(4):195 – 215, October 1986.

[82] D. E. Knuth. *Axioms and Hulls*, volume 606 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, 1992.

[83] Donald Knuth. *The Art of Computer Programming: Vol. 4 Fascicle 4A: Generating All Trees*. Addison-Wesley, 2006.

[84] Johann Linhart and Ronald Ortner. On the combinatorial structure of arrangements of oriented pseudocircles. *The Electronic Journal of Combinatorics*, 11(1), 2004. Article R30 (online).

[85] Mario Livio. *The Equation That Couldn't Be Solved: How Mathematical Genius Discovered the Language of Symmetry*. Simon & Schuster, 2005.

[86] Raymon Llull. *Ars Magna*. Lyons, 1305. In Latin.

[87] Arthur L. Loeb. *Color and Symmetry*. Wiley-Interscience, 1971.

[88] B. B. Mandelbrot. *The Fractal Geometry of Nature*. W.H. Freeman and Company, 1982.

[89] Robert L. Martin. *Studies in Feedback-Shift-Register Synthesis of Sequential Machines*. M.I.T. Press, Cambridge, Massachusetts, 1969.

[90] Trenchard More, Jr. On the construction of Venn diagrams. *Journal of Symbolic Logic*, 24:303–304, 1959.

[91] Frank Morgan. *Geometric Measure Theory: A Beginner's Guide*. Academic Press, second edition, 1995.

[92] Przemysław Nowicki. Koniczynka n-listna [n-leaf clover]. *Wiadomości Matem. (2)*, 19(1):11–18, 1975. In Polish.

[93] Lewis Pakula. A note on Venn diagrams. *Amer. Math. Monthly*, 96(1):38–39, 1989.

[94] C.S. Peirce. *Collected Papers*. Harvard University Press, Cambridge, MA, 1933.

[95] Tony Phillips. Topology of Venn Diagrams. `http://www.ams.org/featurecolumn/archive/venn.html`, 2007. [Online; accessed 08-December-2007].

[96] Vern S. Poythress and Hugo S. Sun. A method to construct convex, connected Venn diagrams for any finite number of sets. *The Pentagon*, 31:80–83, Spring 1972.

[97] Anthony Quas, 2007. Personal communication.

[98] Bill Rankin. Projection Reference. `http://www.radicalcartography.net/?projectionref`, 2006. [Online; accessed 31-August-2007].

[99] Elmer G. Rees. *Notes on Geometry*. Springer-Verlag, Berlin, 1983.

[100] A. Rényi, V. Rényi, and J. Surányi. Sur l'indépendance des domaines simples dans l'Espace Euclidien a n dimensions. *Colloquium Mathematicum*, 2:130–135, 1951. In French.

[101] D. Roberts. *The Existential Graphs of Charles S. Peirce*. Mouton, The Hague, 1973.

[102] P. Rodgers, editor. *Proceedings of the First International Workshop on Euler Diagrams (Euler 2004)*, volume 134 of *Electronic Notes in Theoretical Computer Science*, Amsterdam, 2005. Elsevier Science B.V. Held at the University of Brighton, Brighton, September 22–23, 2004.

[103] Richard L. Roth. Color symmetry and group theory. *Discrete Mathematics*, 38:273–296, 1982.

[104] Todd Rowland. "Spherical Distance." From MathWorld–A Wolfram Web Resource, created by Eric W. Weisstein. `http://mathworld.wolfram.com/SphericalDistance.html`, 2008. [Online; accessed 10-September-2008].

[105] F. Ruskey, C. D. Savage, and T. M. Wang. Generating necklaces. *Journal of Algorithms*, 13:414–430, 1992.

[106] F. Ruskey and M. Weston. A survey of Venn diagrams. *The Electronic Journal of Combinatorics*, 1997. Dynamic survey, Article DS5 (online). Revised 2001, 2005.

[107] F. Ruskey and M. Weston. More fun with symmetric Venn diagrams. *Theory of Computing Systems*, 39(3):413–423, June 2006. Also availabe in *Proceedings of Third International Conference on FUN with Algorithms*, 2004.

[108] Frank Ruskey, Carla Savage, and Stan Wagon. The search for simple symmetric Venn diagrams. *Notices of the American Mathematical Society*, 53(11):1304–1312, December 2006. Includes front cover illustration.

[109] Doris Schattschneider. The plane symmetry groups: Their recognition and notation. *The American Mathematical Monthly*, 85(6):439–450, June–July 1978.

[110] A. Schönflies. *Theorie der Kristallstruktur*. Borntraeger, Berlin, 1923. In German.

[111] R. L. E. Schwarzenberger. *n-Dimensional Crystallography*. Pitman Advanced Publishing Program, 1980.

[112] R. L. E. Schwarzenberger. Colour symmetry. *Bull. London Math. Soc.*, 16(3):209–240, 1984.

[113] Allen J. Schwenk. Venn diagram for five sets. *Mathematics Magazine*, 57(5):297, November 1984.

[114] Marjorie Senechal. Finding the finite groups of symmetries of the sphere. *Amer. Math. Monthly*, 97(4):329–335, 1990.

[115] Micha Sharir. On k-sets in arrangements of curves and surfaces. *Discrete and Computational Geometry*, 6(1):593–613, December 1991.

[116] Sun-Joo Shin. A situation-theoretic account of valid reasoning with Venn diagrams. In J. Barwise et al., editor, *Situation Theory and Its Applications*, volume 2. Stanford: CSLI, 1991.

[117] Sun-Joo Shin. *The Logical Status of Diagrams*. Cambridge University Press, Cambridge, 1994.

[118] Sun-Joo Shin and Oliver Lemon. Diagrams. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2007. `http://plato.stanford.edu/archives/spr2007/entries/diagrams/` [Online; accessed 08-November-2007].

[119] N. J. A. Sloane. The On-Line Encyclopedia of Integer Sequences. `http://www.research.att.com/~njas/sequences/`, 2007. [Online; accessed 8-April-2008].

[120] John P. Snyder. *Map Projections: a Working Manual*. United States Government Printing Office, Washington, D.C, 1987.

[121] Emanuel Sperner. Ein Satz über Untermengen einer endlichen Menge. *Math. Z.*, 27(1):544–548, 1928. In German.

[122] K. Stenning and O. Lemon. Aligning logical and psychological perspectives on diagrammatic reasoning. *Artificial Intelligence Review*, 15(1-2):29–62, 2001. Reprinted in Thinking with Diagrams, Kluwer, 2001.

[123] William T. Trotter. *Combinatorics and Partially Ordered Sets: Dimension Theory*. John Hopkins University Press, 1992.

[124] William T. Trotter. Partially Ordered Sets. In R. L. Graham, M. Grötschel, and L. Lovász, editors, *Handbook of Combinatorics*, volume 1, chapter 8, pages 433 – 481. Elsevier, Amsterdam, 1995.

[125] John Venn. On the diagrammatic and mechanical representation of propositions and reasonings. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 9:1–18, 1880.

[126] John Venn. *Symbolic Logic*. MacMillan, London, 1881. 2nd edition (1894).

[127] Stan Wagon and Peter Webb. Notes: Venn symmetry and prime numbers: a seductive proof revisited. *American Mathematical Monthly*, August/September 2008.

[128] Eric W. Weisstein. “Poulet Number.” From MathWorld–A Wolfram Web Resource. `http://mathworld.wolfram.com/PouletNumber.html`, 2007. [Online; accessed 5-November-2007].

[129] Eric W. Weisstein. “Inversion.” From MathWorld–A Wolfram Web Resource. `http://mathworld.wolfram.com/Inversion.html`, 2008. [Online; accessed 9-October-2008].

[130] Douglas B. West. *Introduction to Graph Theory*. Prentice-Hall, second edition, 2001.

[131] M. Weston. On symmetry in Venn diagrams and independent families. Master's thesis, University of Victoria, 2003.

[132] D. E. White and S. G. Williamson. Recursive matching algorithms and linear orders on the subset lattice. *J. Combinatorial Theory*, 23(2):117–127, 1977.

[133] Wikipedia. List of planar symmetry groups — Wikipedia, The Free Encyclopedia, 2008. [Online; accessed 16-September-2008].

[134] Peter Winkler. Venn diagrams: some observations and an open problem. In *Proceedings of the fifteenth Southeastern conference on combinatorics, graph theory and computing (Baton Rouge, La., 1984)*, volume 45, pages 267–274, 1984.