The Eightfold Way

The Beauty of Klein's Quartic Curve

Silvio Levy
Editor

Cambridge University Press
978-0-521-00419-0 - The Eightfold Way: The Beauty of Klein's Quartic Curve
Edited by Silvio Levy
Frontmatter
More information

Felix Klein discovered in the 1870s that the simple equation $x^3y+y^3z+z^3x=0$ (in complex projective coordinates) describes a surface having many remarkable properties, including 336-fold symmetry – the maximum possible for any surface of this genus. Since then this object has come up in different guises in several areas of mathematics.

The mathematical sculptor Helaman Ferguson has tried to distill some of the beauty and remarkable properties of this surface in the form of a sculpture that he entitled *The Eightfold Way*, permanently installed at the Mathematical Sciences Research Institute in Berkeley.

This volume seeks to explore the rich tangle of properties and theories surrounding this object, as well as its esthetic aspects. It contains:

- The text written by William Thurston to explain the sculpture to a wide public at the time of its inauguration.
- A broad overview of the position of the Klein quartic in mathematics, with articles by Hermann Karcher and Matthias Weber (geometry), Noam Elkies (number theory), and A. Murray Macbeath (Riemann surfaces).
- A historical overview by Jeremy Gray.
- A richly illustrated essay by the sculptor, Helaman Ferguson.
- An exploration of related curves by Allan Adler, with new results and exposition of old ones.
- The first English translation of Klein's seminal article, "On the order-seven transformation of elliptic functions."

Mathematical Sciences Research Institute
Publications

35

The Eightfold Way

Mathematical Sciences Research Institute Publications

Vol. 1	Freed and Uhlenbeck: *Instantons and Four-Manifolds*, second edition
Vol. 2	Chern (editor): *Seminar on Nonlinear Partial Differential Equations*
Vol. 3	Lepowsky, Mandelstam, and Singer (editors): *Vertex Operators in Mathematics and Physics*
Vol. 4	Kac (editor): *Infinite Dimensional Groups with Applications*
Vol. 5	Blackadar: *K-Theory for Operator Algebras*, second edition
Vol. 6	Moore (editor): *Group Representations, Ergodic Theory, Operator Algebras, and Mathematical Physics*
Vol. 7	Chorin and Majda (editors): *Wave Motion: Theory, Modelling, and Computation*
Vol. 8	Gersten (editor): *Essays in Group Theory*
Vol. 9	Moore and Schochet: *Global Analysis on Foliated Spaces*
Vol. 10–11	Drasin, Earle, Gehring, Kra, and Marden (editors): *Holomorphic Functions and Moduli I, II*
Vol. 12–13	Ni, Peletier, and Serrin (editors): *Nonlinear Diffusion Equations and Their Equilibrium States I, II*
Vol. 14	Goodman, de la Harpe, and Jones: *Coxeter Graphs and Towers of Algebras*
Vol. 15	Hochster, Huneke, and Sally (editors): *Commutative Algebra*
Vol. 16	Ihara, Ribet, and Serre (editors): *Galois Groups over* $\mathbb{Q}$
Vol. 17	Concus, Finn, and Hoffman (editors): *Geometric Analysis and Computer Graphics*
Vol. 18	Bryant, Chern, Gardner, Goldschmidt, and Griffiths: *Exterior Differential Systems*
Vol. 19	Alperin (editor): *Arboreal Group Theory*
Vol. 20	Dazord and Weinstein (editors): *Symplectic Geometry, Groupoids, and Integrable Systems*
Vol. 21	Moschovakis (editor): *Logic from Computer Science*
Vol. 22	Ratiu (editor): *The Geometry of Hamiltonian Systems*
Vol. 23	Baumslag and Miller (editors): *Algorithms and Classification in Combinatorial Group Theory*
Vol. 24	Montgomery and Small (editors): *Noncommutative Rings*
Vol. 25	Akbulut and King: *Topology of Real Algebraic Sets*
Vol. 26	Judah, Just, and Woodin (editors): *Set Theory of the Continuum*
Vol. 27	Carlsson, Cohen, Hsiang, and Jones (editors): *Algebraic Topology and Its Applications*
Vol. 28	Clemens and Kollar (editors): *Current Topics in Complex Algebraic Geometry*
Vol. 29	Nowakowski (editor): *Games of No Chance*
Vol. 30	Grove and Petersen (editors): *Comparison Geometry*
Vol. 31	Levy (editor): *Flavors of Geometry*
Vol. 32	Cecil and Chern (editors): *Tight and Taut Submanifolds*
Vol. 33	Axler, McCarthy, and Sarason (editors): *Holomorphic Spaces*
Vol. 34	Ball and Milman (editors): *Convex Geometric Analysis*
Vol. 35	Levy (editor): *The Eightfold Way*
Vol. 36	Gavosto, Krantz, and McCallum (editors): *Contemporary Issues in Mathematics Education*
Vol. 37	Schneider and Siu (editors): *Several Complex Variables*
Vol. 38	Billera, Björner, Green, Simion, and Stanley (editors): *New Perspectives in Geometric Combinatorics*
Vol. 39	Haskell, Pillay, and Steinhorn (editors): *Model Theory, Algebra, and Geometry*
Vol. 40	Bleher and Its (editors): *Random Matrix Models and Their Applications*

Volumes 1–4 and 6–27 are available from Springer-Verlag

The Eightfold Way

The Beauty of Klein's Quartic Curve

Edited by

Silvio Levy

Mathematical Sciences Research Institute

CAMBRIDGE
UNIVERSITY PRESS

Silvio Levy
Mathematical Sciences
Research Institute
1000 Centennial Drive
Berkeley, CA 94720
United States

Mathematical Sciences
Research Institute
1000 Centennial Drive
Berkeley, CA 94720
United States

The Mathematical Sciences Research Institute wishes to acknowledge support by the National Science Foundation.

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 1999

First paperback edition 2001

Printed in the United States of America

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging in Publication data:

The Eightfold way: the beauty of Klein's quartic curve / edited by Silvio Levy.
x, 331 p. : ill. ; 25 cm. (Mathematical Sciences Research Institute publications ; 35)
Includes bibliographical references.
ISBN 0-521-80209-1 (hardback) 0-521-00419-5 (paperback)
1. Curves, Quartic. 2. Klein, Felix, 1849–1925. I. Levy, Silvio.

QA567.2.Q35 1999 98-051730

ISBN 0 521 00419 5 paper

Contents

The Eightfold Way
MSRI Publications
Volume **35**, 1998

MSRI and the Klein Quartic

On November 14, 1993, a marble and serpentine sculpture was unveiled at the Mathematical Sciences Research Institute in Berkeley, an event that marked one of the ways in which MSRI has been reaching out beyond its traditional role. The work had been commissioned from the famous mathematical sculptor Helaman Ferguson, thanks to a generous donation from Mitsubishi Electric Research Laboratories (MERL) made for the purpose. This sculpture, and the mathematical object that lies behind it, are the subject of this book.

Felix Klein discovered in 1878 that a certain surface, whose equation (in complex projective coordinates) he gave very simply as $x^3y + y^3z + z^3x = 0$, has a number of remarkable properties, including an incredible 336-fold symmetry. He arrived at it as a quotient of the upper complex half-plane by a modular group — the group of fractional linear transformations whose coefficients are integers and that reduce to the identity modulo 7. Since then, the same structure has come up in different guises in many areas of mathematics.

Ferguson's sculpture, *The Eightfold Way*, is a distillation of the beauty and remarkable properties of the Klein quartic. (See Plate 1 following page 150.) At the base is a two-color stone mosaic, representing the uniformization of the surface: a regular hyperbolic tesselation shown in the Poincaré model. Rising out of the central tile, a seven-sided black column cups the artist's Carrara marble rendition of the surface, which highlights its tetrahedral symmetry. The name *The Eightfold Way* is explained by the ridges and grooves that crisscross the otherwise smooth hand-polished surface: they represent the same tesselation, after the surface has folded over itself. If you run your finger along these curves, alternating left and right turns at each corner, you always come back to the beginning after eight turns. In the words of Claire Ferguson, the overall effect is that of "...a symphony of elegant counterpoint — as if Gothic tracery and Alhambra tilings were united in one work."

* * *

This book was a long time in the making, and I owe a debt of gratitude to all the contributors, both for their writing and for their good-humored cooperation during the often hectic process of proof review. I particularly want to thank Hermann Karcher and Matthias Weber, who contributed very early, for their patience; Helaman and Claire Ferguson, for supplying slides and for not biting my head off when I proceeded to lose them; Murray Macbeath, for responding promptly to a late request for a contribution; Jeremy Gray, for allowing me to reprint his *Intelligencer* article and for invaluable advice on the Klein translation; Noam Elkies and Allan Adler, for their thoroughness; and Bill Thurston, for putting all of this into motion. Thanks also to Lauren Cowles and Catherine Felgar, respectively Mathematics Editor and Production Editor at Cambridge University Press, for getting the book out in record time.

Silvio Levy
Berkeley, summer and fall 1998

The Eightfold Way
MSRI Publications
Volume **35**, 1998

The Eightfold Way: A Mathematical Sculpture by Helaman Ferguson

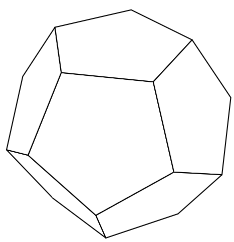
WILLIAM P. THURSTON

This introduction to *The Eightfold Way* and the Klein quartic was written for the sculpture's inauguration. On that occasion it was distributed, together with the illustration on Plate 2, to a public that included not only mathematicians but many friends of MSRI and other people with an interest in mathematics. Thurston was the Director of MSRI from 1992 to 1997.

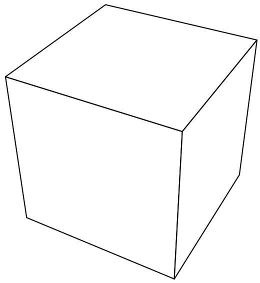
Mathematics is full of amazing beauty, yet the beauty of mathematics is far removed from most people's everyday experience. The Mathematical Sciences Research Institute is committed to the search for ways to convey the beauty and spirit of mathematics beyond the circles of professional mathematicians.

As a step in this effort, MSRI (pronounced "Emissary") has installed a first mathematical sculpture, *The Eightfold Way*, by Helaman Ferguson. The sculpture represents a beautiful mathematical construction that has been studied by mathematicians for more than a century, from many points of view: geometry, symmetry, group theory, algebraic geometry, topology, number theory, complex analysis. The surface depicted by the sculpture was discovered, along with many of its amazing properties, by the German mathematician Felix Klein in 1879, and is often referred to as the Klein quartic or the Klein curve in his honor.

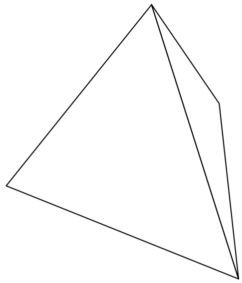
The abstract surface is impossible to render exactly in three-dimensional space, so the sculpture should be thought of as a kind of topological sketch. Ridges and valleys carved into the white marble surface divide it into 24 regions. Each region has 7 sides, and represents the ideal of a regular heptagon (7-gon). The 24 heptagons fit together in triples at 56 vertices. It is the pattern of the division of the surface into heptagons that carries the essence of the mathematics. The Klein quartic thus is an extension of the concept of a regular polyhedron, of which the dodecahedron, the cube and the tetrahedron are examples:



Dodecahedron



Cube



Tetrahedron

Even though the heptagons on the physical surface are not regular, the *pattern* of heptagons on the surface is completely symmetric — in fact, the pattern is just as symmetric as the pattern of pentagons on a dodecahedron. One way to get a sense of the symmetry is to place a finger on any edge. Trace out along the edge to the next intersection, and turn left. Now proceed to the next intersection and turn right. Continue in this way, making a total of 8 turns, LRLRLRLR. If you do this carefully, with concentration and contortion, you arrive back where you started. It doesn't matter where you start or in which direction you go: in 8 alternating turns, you always arrive back at the beginning. (Question: what happens when you do this on a tetrahedron, cube, or dodecahedron?)

In the pattern of heptagons on the surface, and of the 24 heptagons is equivalent to any other heptagon. Furthermore, if any heptagon is rotated by $\frac{1}{7}$th of a revolution, it still fits into the pattern in an identical way. This makes $24 \times 7 = 168$ ways that the pattern of heptagons on the surface can be mapped to itself. Mathematically, the pattern has order 168. When a heptagon is reflected along any of its altitudes, it still fits into the pattern in an identical way, making a total of 336-fold symmetry when the mirror-image transformations are allowed.

The circular base area of the sculpture is also tiled by heptagonal tiles, in a regular geometric pattern that resembles a honeycomb. The sides of the heptagons are arcs of circles; when these arcs are continued, they meet the boundary at a $90°$ angle. This circular base area is a map of the hyperbolic or non-Euclidean plane. In hyperbolic geometry, it is possible to construct regular heptagons whose angles are exactly $120°$; these heptagons fit together to tile the hyperbolic plane. The physical map of the hyperbolic plane is distorted, but in hyperbolic geometry itself all the heptagons have an identical size and shape.

The heptagonal tiling of the base and the heptagonal tiling of the surface are closely related. The 7-sided column that supports the sculpture starts off this relationship: it sweeps up from the central heptagon in the hyperbolic plane to one of the 24 heptagons on the surface. Imagine continuing this relationship. The 7 heptagons adjacent to the foot of the column sweep up and stretch to cover the 7 heptagons that border the top of the column, and so forth. The base area stretches out and wraps around the surface to completely encompass it; it continues stretching and wrapping around and around, infinitely often.

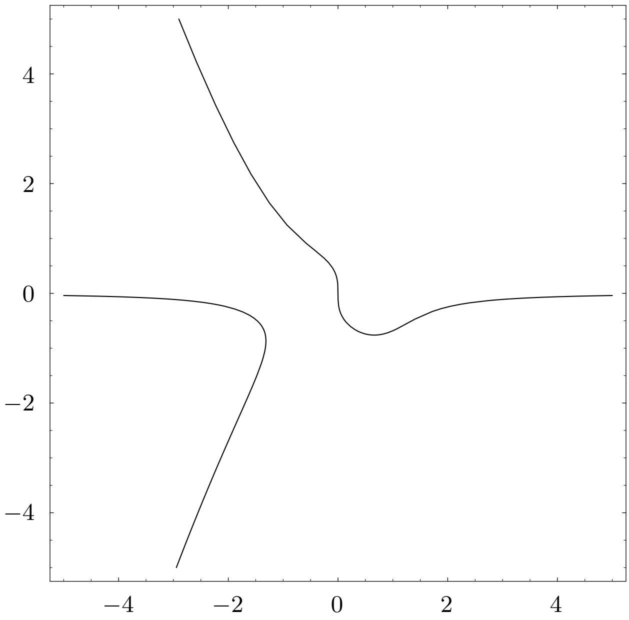
Plate 2 following page 150 shows the heptagonal hyperbolic honeycomb with a pattern superimposed to indicate what happens when it wraps around the surface. The infinite hyperbolic honeycomb is divided into 3 kinds of groups of 8 cells each, where each group is composed of a heptagon together with its 7 neighbors. There are red rings surrounding one person, green groups surrounding another person, and white groups with letters.

When the honeycomb is wrapped around the surface, equivalent groups wrap up to the same place on the surface. In other words, the pattern superimposed on the surface would have only one green group, one red group, and one white group, making 24 heptagons in all. You can check this out by testing the LRLRLRLR rule on the hyperbolic honeycomb. For instance, if you start on an edge that points in toward the central white area and has a red group on its left and a green group on its right, and proceed LRLRLRLR, you will arrive at another edge with red on its right and green on its left. If the initial edge pointed toward and "a" (say), the final edge also points toward an "a".

It is interesting to watch what happens when you rotate the pattern by a $\frac{1}{7}$ revolution about the central tile: red groups go to red groups, green groups go to green groups and white groups go to white groups. The person in the center of a green group rotates by $\frac{2}{7}$ revolution, and the person in the center of a red group rotates by $\frac{4}{7}$ revolution. The interpretation on the surface is that the 24 cells are grouped into 8 affinity groups of 3 each. The symmetries of the surface always take affinity groups to affinity groups. This is analogous to the dodecahedron, whose twelve pentagonal faces are divided into 6 affinity groups of 2 each, consisting of pairs of opposite faces.

The name "Klein quartic" or "Klein curve" refers to an algebraic description of the ideal surface that the sculpture represents, determined by the equation $x^3y + y^3 + x = 0$. (This equation is called a quartic or 4th-degree equation because the highest term x^3y has 3 x's and 1 y, making degree 4 in all.) The solutions to this equation in the (x, y)-plane form the curve shown at the top of the next page. [A more symmetric view is presented in Figure 10 on page 326. –Ed.]

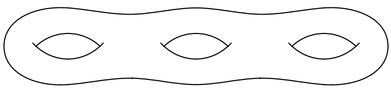
But when x and y are allowed to be complex numbers, there are many more solutions; in fact, the set of solutions forms a 2-dimensional surface in 4-dimensional space. The symmetry of the surface is reflected algebraically by the phenomenon that there are many possible substitutions that keep the equation the same. For instance, if you replace x by Y/X and y by $1/X$, the equation becomes $Y^3/X^4 + 1/X^3 + Y/X = 0$; if you multiply both sides by X^4 to clear denominators, you get the original equation. There are 168 essentially different algebraic "substitutions" that preserve the equation, one for each of the orientation-preserving symmetries of the surface. (Coordinates can be chosen so that the



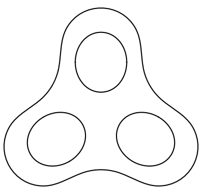
$x^3y + y^3 + x = 0$

center of the central tile of the hyperbolic honeycomb maps to the solution $(0, 0)$ to the equation, and a vertical line through that point maps to the curve graphed above. The substitution just described takes the integral sign to the person in a red ring, the person in a red ring to the person in a green ring, and the person in the green ring back to the integral sign.) Most of the substitutions are more complicated, involving complex algebraic numbers, and we won't describe them.

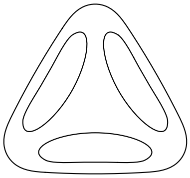
Topologically, the Klein curve is called a surface of genus 3 or a 3-holed torus. Why 3? Here is a standard picture of a 3-holed torus:



In topology, two figures are equivalent if one can be deformed into the other. So we can rearrange the holes (stretching but not tearing or gluing) however we like without changing the topology, for instance into



(to reveal a kind of 3-fold symmetry that was not evident before), and further into



which looks like the frame of a tetrahedron as seen from above. This is the approximate form of the sculpture, and it displays the maximal amount of the symmetry of the ideal surface that can be made directly visible in space.

The heptagonal hyperbolic honeycomb has an interesting relationship to the Fibonacci sequence

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \ldots,$$

in which each number, starting with the third, is the sum of the preceding two. Imagine growing the hyperbolic honeycomb like a crystal, starting with the central white group of 8 heptagons as a seed. At each unit of time, adjoin a heptagon wherever there is a concave angle—that is, adjoin all the heptagons that touch at least two of the heptagons already present. At the first step, you will add 7 heptagons. Second, you will add the 14 green heptagons that fill out the next complete ring. On the third step, you add 21 heptagons, consisting of 14 red and 7 white heptagons (the centers of green groups). The sequence, if we include the ring of 7 white heptagons in the initial seed, goes

$$7, 7, 14, 21, 35, 56, 91, 147, 238, 385, 623, \ldots.$$

Each term is the sum of the preceding two: this sequence is 7 times the Fibonacci sequence!

The number of tiles grows very rapidly as you add additional layers. That is why the tiles around the edges must get quite small in the map of the hyperbolic plane: there are so many of them that otherwise they wouldn't fit. The base of the sculpture includes tiles corresponding to the first 7 terms of the sequence, making 231 tiles in all (or 232 if you include the spot where the column fits). The cover diagram shows the tiles for two additional terms plus a few scattered heptagons, making over 617.

Instructions for how to glue the 24 heptagons of the surface together can be constructed as follows. Label the 24 heptagons with different labels, say the letters `a` through `x`. Arrange these letters, together with a sharp sign (`#`), in a grid pattern in the plane that repeats very 7 units across, every 7 units up, and is symmetric about each `#`. All the information is contained in a 7×7 table. Notice that the 7×7 table is filled out by the `#` together with 2 occurrences of each of the 24 letters:

#	a	b	c	c	b	a	#	a	b	c	c	b	a	#
d	e	f	g	h	i	j	d	e	f	g	h	i	j	d
k	l	m	n	o	p	q	k	l	m	n	o	p	q	k
r	s	t	u	v	w	x	r	s	t	u	v	w	x	r
r	x	w	v	u	t	s	r	x	w	v	u	t	s	r
k	q	p	o	n	m	l	k	q	p	o	n	m	l	k
d	j	i	h	g	f	e	d	j	i	h	g	f	e	d
#	a	b	c	c	b	a	#	a	b	c	c	b	a	#
d	e	f	g	h	i	j	d	e	f	g	h	i	j	d
k	l	m	n	o	p	q	k	l	m	n	o	p	q	k
r	s	t	u	v	w	x	r	s	t	u	v	w	x	r
r	x	w	v	u	t	s	r	x	w	v	u	t	s	r
k	q	p	o	n	m	l	k	q	p	o	n	m	l	k
d	j	i	h	g	f	e	d	j	i	h	g	f	e	d
#	a	b	c	c	b	a	#	a	b	c	c	b	a	#
d	e	f	g	h	i	j	d	e	f	g	h	i	j	d
k	l	m	n	o	p	q	k	l	m	n	o	p	q	k
r	s	t	u	v	w	x	r	s	t	u	v	w	x	r
r	x	w	v	u	t	s	r	x	w	v	u	t	s	r
k	q	p	o	n	m	l	k	q	p	o	n	m	l	k
d	j	i	h	g	f	e	d	j	i	h	g	f	e	d
#	a	b	c	c	b	a	#	a	b	c	c	b	a	#

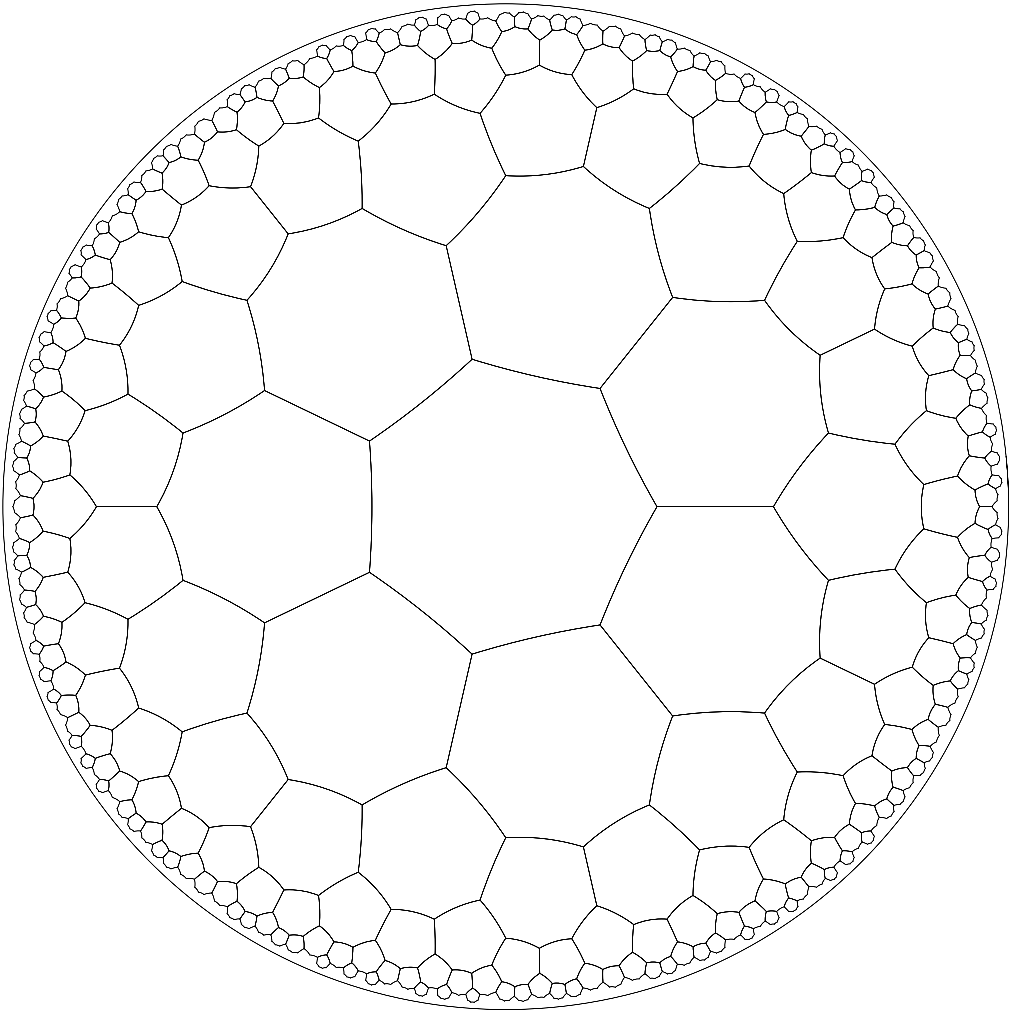
To determine what heptagons to glue to a given heptagon (call it `z`), find the letter of the heptagon in the table. It's always possible to construct a line segment that connects some `#` to `z` without going through any intermediate letters. Draw a line parallel to `#z` that is as close as possible to the right while still going through letters in the table. The letters along this line are the heptagons adjacent to `z`, in counterclockwise order. For example, the `a` heptagon is glued to the 7 heptagons in the second row of the table: `defghij`. The `e` heptagon is glued to `adltvnf`. The neighbors of `t` are slightly harder to determine: `t` is 2 units to the left and 3 units up from a `#`, so starting with an `l`, which is 2 units to the left and 2 units up, one can repeatedly go 2 over and 3 up, reading off `loirbve`.

Try labeling the blank heptagonal honeycomb on the next page, using this rule.

The spirit of mathematics and the essence of its beauty is remarkably fragile, because mathematics is about ideas and about thought. Mathematics takes place in the mind, and no two minds are the same. After many years of study and work, a mathematician may stumble on a vast and beautiful vista that unifies and simplifies many things that once appeared disparate and complicated. Mathematicians can share a beautiful mathematical vista with one another, but there is no camera that can easily capture an image of such a vista to convey it in full to people who have not trudged along many of the same trails.

We have only touched on a small part of the mathematical vista associated with this sculpture, but we hope that you can get form it some glimpse of the unity, the beauty, and the spirit of mathematics.

William P. Thurston
University of California Davis
Department of Mathematics
565 Kerr Hall
Davis, CA 95616
United States
wpt@math.ucdavis.edu



The Eightfold Way
MSRI Publications
Volume **35**, 1998

The Geometry of Klein's Riemann Surface

HERMANN KARCHER AND MATTHIAS WEBER

Abstract. Starting from the hyperbolic definition of Klein's surface we prove platonicity, derive the two classical equations $W^7 = Z(Z-1)^2$ between meromorphic functions and $x^3y+y^3z+z^3x = 0$ between holomorphic forms, describe a pair of pants decomposition in terms of which the automorphisms can be seen, find a basis for the holomorphic forms for which all periods (hence the Jacobian) can be computed, and from which the lattice of rhombic tori can be determined which are covered by Klein's surface.

1. Introduction

In autumn 1993, in front of the MSRI in Berkeley, a marble sculpture by Helaman Ferguson called *The Eightfold Way* was revealed. This sculpture shows a compact Riemann surface of genus 3 with tetrahedral symmetry and with a tessellation by 24 distorted heptagons. The base of the sculpture is a disc tessellated by hyperbolic 120°-heptagons, thus suggesting that one should imagine that the surface is "really" tessellated by these regular hyperbolic polygons. In the celebration speech Bill Thurston explained how to see the surface as a hyperbolic analogue of the Platonic solids: Its symmetry group is so large that any symmetry of each of the 24 regular heptagons extends to a symmetry of the whole surface — a fact that can be checked "by hand" in front of the model: Extend any symmetry to the neighboring heptagons, continue along arbitrary paths and find that the continuation is independent of the chosen path. The hyperbolic description was already given by Felix Klein after whom the surface is named. The large number of symmetries — we just mentioned a group of order $24 \cdot 7 = 168$ — later turned out to be maximal: Hurwitz [1893] showed that a compact Riemann surface of genus $g \geq 2$ has at most $84(g-1)$ automorphisms and the same number of antiautomorphisms. The next surface where Hurwitz's bound is sharp is treated in [1965]; see also [Lehner and Newman 1967; Kulkarni 1982], as well as Macbeath's article in this volume.

Karcher was partially supported by MSRI and SFB256, and Weber through a one year grant of the DFG.

The sculpture introduced Klein's surface to many non-experts. Of course the question came up how the hyperbolic definition of the surface (as illustrated by the sculpture) could be related to the rather different algebraic descriptions. For example, the equation

$$W^7 = Z(Z-1)^2$$

relates two meromorphic functions on the surface, and

$$x^3y + y^3z + z^3x = 0$$

relates three holomorphic 1-forms. The answer to this question is known in general: The uniformization theorem implies that every Riemann surface of genus $g \geq 2$ has a hyperbolic metric, that is, a metric of constant curvature -1, and the existence of sufficiently many meromorphic functions implies that every compact Riemann surface has an algebraic description. But it is very rare that one can pass *explicitly* from one description to the other.

There were other natural questions. In the hyperbolic picture one sees cyclic automorphism groups of order 2, 3, and 7 — what are the quotient surfaces? Topologically, this can easily be answered with the Euler number of a tessellated surface, $\chi = F - E + V$, if one takes a tessellation that passes to the quotient. Moreover, we will identify the quotient map under the order 7 subgroup with the meromorphic function Z in the first equation. By contrast, we do not know a group theoretic definition of the other function, W; it is constructed in the hyperbolic picture with the help of the Riemann mapping theorem. — The quotient surfaces by the other groups above, those of order 2 and 3, are always tori. This has another known consequence: Klein's surface does not doubly cover the sphere, it is not "hyperelliptic" — but it also leads to more questions: What tori appear as quotients? The differential of a holomorphic map to a torus is a holomorphic 1-form whose period integrals (along arbitrary closed curves on the surface) *are* the lattice of the torus. So again, the question is highly transcendental in general and explicit answers are rare.

Here the answer is possible, since we can identify Klein's surface in yet another representation of compact Riemann surfaces. Consider the Riemann sphere endowed with a flat metric with cone singularities. Riemann surfaces can be described as coverings over such a sphere that are suitably branched over the cone singularities. In this situation one has a developing map from the Riemann surface to the complex plane. Its differential is a holomorphic 1-form on the universal cover whose zeros are at the cone singularities. With a good choice of the flat metric this 1-form actually descends to the compact Riemann surface! (Already this step rarely succeeds.) In the special case of Klein's surface we find with the help of the 7-fold covering mentioned above *three different* such representations. This gives a basis of the holomorphic 1-forms — in fact the forms x, y, z of the second equation above — for which the *periods can be computed* via the Euclidean geometry of the flat metrics. At this point the Jacobian of the surface is determined. We proceed to find linear combinations of the basis 1-forms so that their periods are a lattice in $\mathbb{C}$. This shows that the Jacobian is the product of three times the same rhombic torus with diagonal ratio $\sqrt{7} : 1$. This torus has "complex multiplication", namely we can map its lattice to an index 2 sublattice by multiplication with $\bigl(1+\sqrt{-7}\,\bigr)/2$. This leads to recognizing the lattice as the ring of integers in the quadratic number field $\mathbb{Q}\bigl(\sqrt{-7}\,\bigr)$ and to see that this torus is defined over $\mathbb{Q}$.

We learnt from [Rodríguez and González-Aguilera 1994] that the hyperbolic description of the Fermat quartic,

$$x^4 + y^4 + z^4 = 0,$$

is surprisingly similar to Klein's surface. In fact, *each* Fermat surface is platonically tessellated by π/k-triangles; the area of these tiles is $\pi(k-3)/k$, which is always larger than the area $\pi/7$ of the $2\pi/7$-triangles, which are 56 platonic tiles for Klein's surface. Also, Jacobians and, for $k = 4$, quotient tori can be computed with the methods outlined above. We included this only because we found a comparison instructive. The result is less exciting because the questions above can be answered for the Fermat case in each description separately.

Acknowledgment. This work started from discussions about the Ferguson sculpture while the authors enjoyed the hospitality of the MSRI. Conversations with W. Thurston and M. Wolf were then particularly helpful. Later a large number of colleagues helped us with comments, questions and advice.

This paper is organized as follows:

SECTION 2: Summarizes a few facts from the *group theoretic* treatment of platonic surfaces.

SECTION 3: Treats *two genus* 2 platonic surfaces. Together they show many phenomena that we will also encounter with Klein's surface, but they are much simpler. We hope this will help the reader to see more quickly where we are heading in the discussion of Klein's surface.

SECTION 4: Deduces *Klein's surface* from assumptions that require less than its full symmetry, derives the above equations and proves platonicity.

SECTION 5: Describes a *pairs of pants* decomposition that emphasizes the symmetries of one S_4 subgroup of the automorphism group. These pants also allow to list the conjugacy classes of all automorphisms.

SECTION 6: Discusses and compares the *Fermat surfaces*, in particular the quartic.

SECTION 7: Introduces *flat cone metrics*. In terms of these we construct holomorphic forms with computable periods, determine the Jacobians of the discussed examples and find explicit maps to tori. We prove that all quotient tori of Klein's surface are the same rhombic torus with diagonal ratio $\sqrt{7} : 1$.

2. Triangle Groups and Platonic Surfaces

Platonically tessellated Riemann surfaces and the structure of triangle groups are closely related. To give some background information we summarize the following known facts.

A *symmetry* of a Riemann surface is an isometry with respect to the hyperbolic metric on it. An *automorphism* is an orientation-preserving symmetry. This is the same as a conformal automorphism. Thus we do not mean that a symmetry has to be the symmetry of some embedding (like the sculpture) or immersion of the surface [Schulte and Wills 1985].

A tessellation of a Riemann surface is *platonic* if the symmetry group acts transitively on flags of faces, edges and vertices. Such a tessellation is also called a *regular map* [Coxeter and Moser 1957]. Finally, a Riemann surface is called platonic if it has some platonic tessellation.

Suppose now that we have a Riemann surface M^2 that is platonically tessellated by regular k-gons with angle $2\pi/l$. The stabilizer of one polygon in the symmetry group of the surface then contains at least the dihedral group of the polygon. Consequently there is a subgroup of the symmetry group that has as a fundamental domain a hyperbolic triangle with angles $\pi/2$, π/k, π/l. We will call such triangles from now on $(2,\, k,\, l)$-triangles. Observe that the order of this group is

$$\begin{aligned}\text{order} &= \text{hyperbolic area}(M^2)\big/\text{area}(2,\, k,\, l)\text{-triangle}\\ &= -2\pi\cdot\chi(M^2)\Big/\left(\frac{\pi}{2}-\frac{\pi}{k}-\frac{\pi}{l}\right),\end{aligned}$$

and half as many automorphisms are orientation-preserving. The smallest possible areas of such triangles are

$$\text{area}(2,3,7)=\pi/42\,,\quad \text{area}(2,3,8)=\pi/24\,,\quad \text{area}(2,4,5)=\pi/20\,.$$

Now consider the group generated by the reflections in the edges of a $(2,\, k,\, l)$-triangle in the hyperbolic plane; this group is called a triangle group. It acts simply transitively on the set of triangles. The covering map from $\mathbb{H}^2$ to M^2 maps triangles to triangles; the preimage of each triangle defines the classes of equivalent triangles in $\mathbb{H}^2$. The deck group of M^2 acts simply transitively on each equivalence class and, because we assumed M^2 to be platonically tessellated, it is also true that the (anti-)automorphism group of M^2 acts simply transitively on the set of equivalence classes. This shows that the deck group of the surface is a (fixed point free) normal subgroup of the triangle group.

Vice versa, given a fixed point free normal subgroup N of a $(2,\, k,\, l)$-triangle group G, then we define a Riemann surface M^2 as the quotient of $\mathbb{H}^2$ by N. This surface is tessellated by the $(2,\, k,\, l)$-triangles and the factor group G/N acts simply transitively on these triangles. In $\mathbb{H}^2$ the $(2,\, k,\, l)$-triangles of course fit together to a pair of dual platonic tessellations, one by k-gons with angle $2\pi/l$, the other by l-gons with angle $2\pi/k$. Both tessellations descend to tessellations of the quotient surface (namely: Consider the projection of the polygon centers in $\mathbb{H}^2$ to the surface, we recover a polygon tessellation of the surface as the Dirichlet cells around the projected set of centers). They are still platonic.

It is therefore in an obvious way equivalent to consider compact platonically tessellated Riemann surfaces or finite index normal subgroups of triangle groups. (We may even allow $(2, k, \infty)$-triangles, noncompact finite area triangles with one 0-angle.) See [Bujalance and Singerman 1985; Singerman 1995].

Meromorphic functions and forms are now accessible from this group theoretic approach as automorphic functions and forms on the hyperbolic plane with respect to the deck group of the surface. With the group-theoretical approach one does not always find the simplest equations [Streit 1996]. By contrast, in our discussion of Klein's surface we will construct on it simple functions and forms for which we do *not* know a group theoretic definition. We use the following two methods:

Meromorphic functions: We map one tile of the tessellated Riemann surface to a suitable spherical domain with the Riemann mapping theorem; we extend this map by reflection across the boundary and finally check that the extension is compatible with the identifications.

Holomorphic 1-forms: We take exterior derivatives of developing maps of flat cone metrics and check by holonomy considerations whether they are well defined on the surface.

3. Two Platonic Surfaces of Genus Two

We explain with the simplest hyperbolic examples how symmetries can be used to derive algebraic equations.

3.1. The $\pi/5$-case. Let's try to construct a genus 2 surface M^2 that is platonically tessellated by F equilateral $\pi/5$-triangles. Such a triangulation must have $E = \frac{3}{2} \cdot F$ edges and $V = \frac{3}{10} \cdot F$ vertices, since 10 triangles meet at a vertex. Euler's formula then gives

$$\chi(M^2) = -2 = F \cdot \left(1 - \frac{3}{2} + \frac{3}{10}\right), \qquad F = 10, \qquad V = 3\,.$$

Equivalently, we could have used the Gauß-Bonnet formula

$$-2\pi \cdot \chi(M^2) = \operatorname{area}(M^2) = F \cdot \operatorname{area}(\text{triangle}) = F \cdot \frac{2\pi}{5}\,.$$

These ten triangles fit around one vertex to form a $2\pi/5$-decagon, which is already a fundamental domain for the surface we want to construct. What remains to be done is to give suitable identifications. We consider only identifications that satisfy necessary conditions for *platonic* tessellations. For example, we want the $2\pi/5$-rotations around the center of the decagon to extend to symmetries of the surface. This implies that the identification of one pair of edges determines all the others. Since the angles at five decagon vertices sum up to 2π the edge identifications have to identify every second vertex. This leaves only two possibilities, which will turn out to define the same surface: Identify edge 1 to edge 6 or to edge 4. Both cases are promising, because further necessary conditions for platonicity are satisfied.

Synthetic arguments in Euclidean and hyperbolic geometry are very much the same: One can compose two reflections in orthogonal lines to obtain a 180° rotation; one can join the centers of two 180° rotations by a geodesic and take it and a perpendicular geodesic through either center as such reflection lines; this shows that the composition of two involutions "translates" the geodesic through their centers. In the hyperbolic case this is the only invariant geodesic, it is also called *axis* of the translation.

Platonicity implies that the midpoints of edges are centers of 180°-rotations. On a compact platonic surface one can therefore extend any geodesic connection of midpoints of edges by applying involutions until one gets a closed geodesic. (Note that these extensions meet the edges, at the involution centers, always with the same angle and there are only finitely many edges.) This means that we always find translations in the deck group that are generated by involutions. Therefore, if we want to construct a platonic surface, then it is a good sign if already the identification translations are products of involutions. This is true for both identification candidates above: For the identification of the opposite edges (say) 1 and 6 take as centers the midpoint of the decagon and the midpoint of edge 6; the translation that identifies edges 1 and 4 is the product of the involutions around the marked midpoints of the radial triangle edges 1 and 3. See Figure 1 for the axes of these translations.

We are now going to construct meromorphic functions on M^2 since this leads to an algebraic definition of the surface. Namely, if two functions have either no common branch points or else at common branch points relatively prime branching orders then they provide near any point holomorphic coordinates — that is, an atlas for the surface. To turn this into a definition one needs to specify the change-of-coordinates and the classical procedure is to do this by giving an algebraic relation between the two functions. Therefore, to describe a specific hyperbolic surface algebraically means that one has to construct two meromorphic functions that one understands so well that one can deduce their algebraic relation. There is no general procedure to achieve this. In highly symmetric situations one can divide by a sufficiently large symmetry group and check whether the quotient Riemann surface is a sphere. Any identification of this quotient sphere with $\mathbb{C} \cup \{\infty\}$ turns the quotient map into a meromorphic function. This method is sufficient for the following genus 2 examples. Another way to construct meromorphic functions is to use the Riemann mapping theorem together with the reflection principle to produce first maps from a fundamental domain of an appropriate group action on the surface to some domain on the sphere and extend this by reflection to a map from the whole surface to the

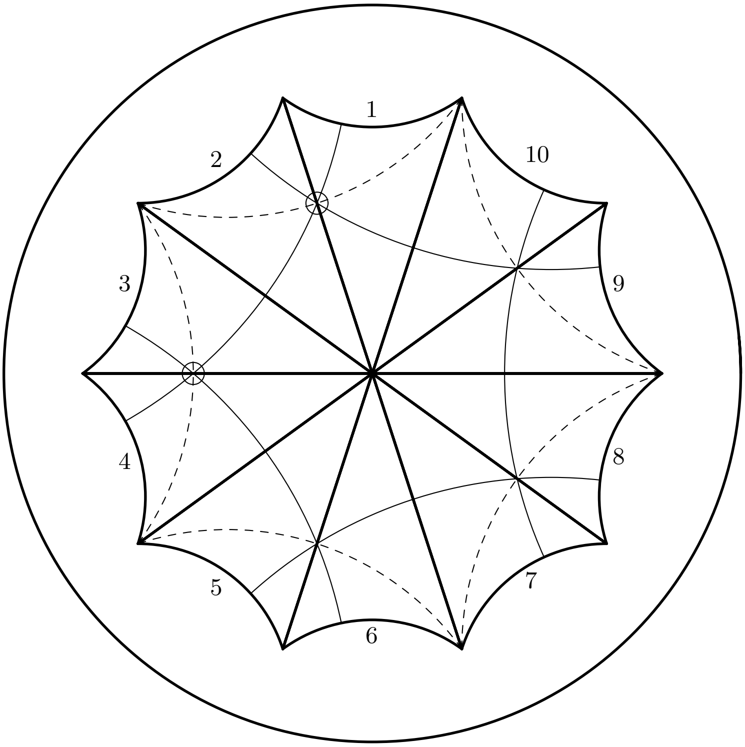


Figure 1. Decagon composed of ten equilateral hyperbolic $\frac{\pi}{5}$-triangles.

sphere. This method will be of importance for Klein's surface, and we will explain it with the simpler functions on the genus 2 surfaces.

To define the first function, look at the order 5 rotation group around the center of the decagon. This group respects the identification and therefore acts on the surface by isometries. It has 3 fixed points, namely the center of the decagon and the two identified sets of vertices. Using Euler's formula we see that the quotient surface is a sphere: Take any triangulation of M^2 that is invariant under the rotation group. Then the quotient surface is also triangulated and denoting by f the number of fixed points of the rotations on M^2, we compute its Euler number

$$\tfrac{1}{5}\big((V-f)-E+F\big)+f \;=\; \tfrac{1}{5}(-2-f)+f \,\in\, \{0,2\},$$

which reproves $f=3$ and shows that $\chi=2$ for the quotient.

This function can also be understood via a Riemann mapping problem: Imagine that the ten triangles are alternatingly colored black and white, "Riemann map" a white triangle to the upper half plane, "Möbius normalize" so that the vertices go to $0,\ 1,\ \infty$, and extend analytically by reflection in the radial boundaries to a map from the decagon to a fivefold covering of the sphere, branched

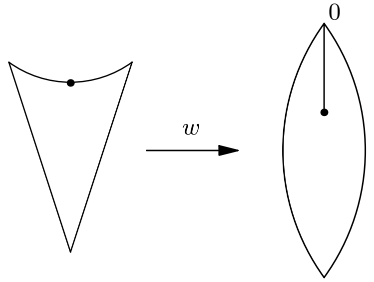


Figure 2. Mapping a hyperbolic $\frac{\pi}{5}$-triangle to a spherical $\frac{2\pi}{5}$-sector.

over ∞ and with slits from 0 to 1 on each sheet. Finally identify the edges of the slits in the same way as the preimage edges of the decagon. Therefore we can either see the quotient sphere as isometric to the double of a hyperbolic $2\pi/5$-triangle, which gives a singular hyperbolic metric on the sphere, or we can see M^2 as a fivefold covering of the Riemann sphere, branched over $0,\ 1,\ \infty$. In any case we have obtained — for both identification patterns — a meromorphic function z on M^2 that sends the three vertices of the triangulation as fivefold branch points to $0,\ 1,\ \infty$.

For a second function, we can consider the quotient of M^2 by any involution to obtain

$$\tfrac{1}{2}\left((V-f)-E+F\right)+f \;=\; \tfrac{1}{2}(-2-f)+f \;\in\; \{0,2\}\,;$$

hence $f=2$ or $f=6$. In both of our cases take as the involution one of those that were used to define the identifications and observe that we have $f=6$ (for the first identification we have as fixed points the midpoint of the decagon and the identified midpoints of opposite edges) so that the quotient by this involution again is a sphere. We normalize this meromorphic quotient function w on M^2 up to scaling by sending the midpoint of the decagon to ∞ and the two other vertices of the triangulation to 0 (and similarly for the other identification pattern).

Since reflection in the radial triangle edges passes to the quotient we can also understand the function w as mapping each triangle to a spherical $2\pi/5$-sector that is bounded by great circle arcs from 0 to ∞ and has a straight slit in the direction of the angle bisector, as in Figure 2. By scaling we may take the slit to have arbitrary length.

Simply by comparing the divisors of z and w we see that w^5 and $z(z-1)$ are proportional functions and (after scaling w) we obtain

$$w^5 = z(z-1),$$

which is a defining equation for M^2, the same for both identification patterns.

We will now be disappointed and find that the triangle tessellation is *not* platonic. One way to see this is to check that the involutions around midpoints of edges that were *not* used to define the identifications are not compatible with them. A more algebraic way is to produce too many holomorphic 1-forms by considering the following divisor table, where we define $y = z/(z-1)$:

vertices	V_1	V_2	V_3
z	0^5	1^5	∞^5
w	0	0	∞^2
y	0^5	∞^5	1^5
dy/y	∞	∞	0^4
$w \cdot (dy/y)$	$\star$	$\star$	0^2
$w^2 \cdot (dy/y)$	0	0	$\star$

Now suppose that M^2 were platonically tessellated. Then the 120°-rotation of one triangle would extend to a symmetry of the whole surface. This implies that we could cyclically permute the divisor of the holomorphic 1-form $w^2 \cdot dy/y$ to get divisors of other forms. The quotient of two of these would be a meromorphic function with only one simple pole, a contradiction.

Fortunately, we have not lost completely since we can platonically tessellate M^2 with two $\pi/5$-pentagons by joining even numbered neighboring vertices of the decagon, dashed in Figure 1. This is not quite as good as hoped for, but also on Klein's surface we will find platonic and other non-platonic tessellations by regular polygons.

3.2. The $\pi/4$-case. Next we will construct a more symmetric platonic genus 2 surface; its automorphism group has order 48, the maximum for genus 2 [Burnside 1911]. The quotient sphere is the double of the hyperbolic $(2,3,8)$-triangle—which is less than twice as big as the doubled $(2,3,7)$-triangle in Klein's case. We want the surface to be platonically tessellated by equilateral $\pi/4$-triangles. Since eight such triangles fit around one vertex we have

$$\chi(M^2) = -2 = F \cdot \left(1 - \tfrac{3}{2} + \tfrac{3}{8}\right), \qquad F = 16, \qquad V = 6\,.$$

The eight triangles around one center vertex form a small $\pi/2$-octagon. The remaining 8 triangles can be placed along the edges. No other pattern would be possible for a platonic surface, because the 45°-rotation around the center vertex extends to a symmetry of the surface. Hence we expect as a fundamental domain of our surface a big regular $\pi/4$-octagon (Figure 3).

Again we try the identification of opposite edges by hyperbolic translations, this time no other candidate is possible by platonicity. As before, these translations are compositions of two 180°-rotations (around the midpoint of a boundary edge and around the center of the octagon, both of which are triangle vertices). Moreover, all vertices (angle $2\pi/8$) of the big octagon are identified to one vertex to give a smooth hyperbolic genus 2 surface M^2. The 180°-rotation around the midpoint of the octagon is an involution of M^2 whose fixed points are the 6 vertices of the 16 $\pi/4$-triangles. The projection z to the quotient goes again to a sphere. One easily checks that this involution commutes with all reflections in the triangle edges so that these reflections and their fixed points pass to the

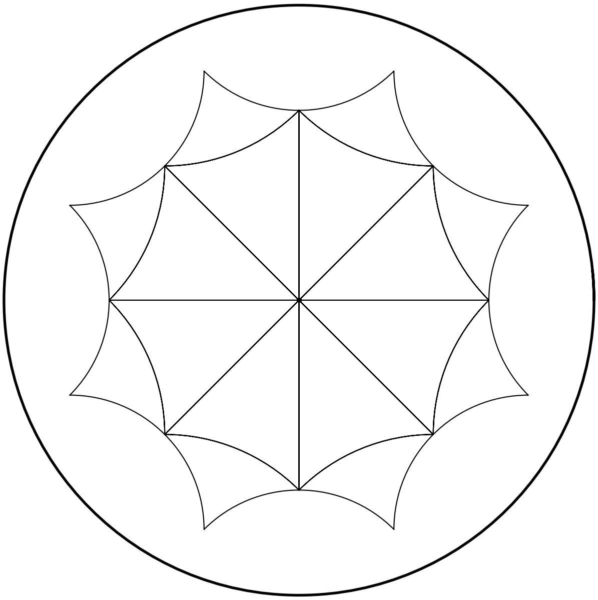


Figure 3. Hyperbolic $\frac{\pi}{4}$-octagon with 16 equilateral triangles.

quotient. Hence the hyperbolic quotient metric on the sphere is necessarily given by the octahedral tessellation by hyperbolic $\pi/4$-triangles. We can assume that the octahedron has its vertices in $0, \pm 1, \pm \mathbf{i}$ and ∞.

As before, this quotient map can also be defined independently: first Riemann-map a hyperbolic $\pi/4$-triangle to a spherical $\pi/2$-triangle, then extend analytically by reflection in the edges and check compatibility with the identifications.

Since M^2 is only a double covering over the sphere with known branch values, we have the following equation for this Riemann surface:

$$w^2 = z \cdot \frac{z-1}{z+1} \cdot \frac{z-\mathbf{i}}{z+\mathbf{i}}.$$

We still have to prove platonicity. Since all the reflections in symmetry lines of the octagon are clearly compatible with the identifications we only have to check that the involution around the midpoint of one radial triangle edge is also compatible. This can be seen by checking in the tessellated hyperbolic plane that *any* two vertices that are two triangle edges apart are equivalent under the identifications. It can also be seen on the doubly covered octahedral tessellation of the sphere by introducing three branch cuts and checking that $180°$-rotation around the midpoint of an octahedron edge on one sheet extends to a symmetry of the double cover. One observes that this involution has only two fixed points since at the antipodal point of the sphere the sheets are interchanged; the quotient map therefore only goes to a torus. Since this involution commutes with a reflection of M^2 the quotient torus has also such a symmetry, called a complex conjugation; the fixed point set of this torus reflection has *two* components — that is, the torus is rectangular. In the case of Klein's surface all the involutions will give quotient maps to rhombic tori.

With platonicity established we can interpret the function w above as the quotient map under the rotations of order 3 around the center of one triangle. These rotations can be seen on the octahedral sphere as follows: Consider a 120°-rotation of the octahedron around the centers of two opposite triangles. This map lifts to an isometry of M^2 with the desired property. It has four fixed points over the two fixed points of the rotation of the octahedron.

Clearly, a fundamental domain for the group of all automorphisms now is one third of one $\pi/4$-triangle, i.e., two (2, 3, 8)-triangles, each of area $\pi/24$. This gives for the order of the automorphism group $-2\pi\chi/(2\pi/24) = 48$. Why is this the maximal order for genus 2? A proof of Hurwitz' theorem begins by dividing a Riemann surface of genus ≥ 2, endowed with its hyperbolic metric, by the full group of automorphisms. These are also hyperbolic isometries. The quotient is a Riemann surface with larger Euler number and a hyperbolic metric with π/k_i cone singularities. The automorphism group is maximal (for the considered genus) if the hyperbolic area of the quotient surface is minimal. The two smallest quotients are the doubles of the hyperbolic (2, 3, 7)- and (2, 3, 8)-triangles. Therefore we have to show that (2, 3, 7) does not occur for genus 2. But already a cyclic group of prime order $p \geq 7$ is impossible for genus 2, since, from the Euler number of the quotient, the possibilities for the number f of fixed points of this group are given by

$$\frac{1}{p} \cdot \big((V - f) - E + F\big) + f \;=\; -\frac{1}{p}(2 + f) + f \;\in\; \{0, 2\},$$

or

$$f \;\in\; \left\{\frac{2}{p-1},\; 2 + \frac{4}{p-1}\right\} \subset \mathbb{Z}\,, \quad \text{with } p \;\in\; \{2, 3, 5\}\,.$$

4. The Hyperbolic Description of Klein's Surface

Klein's surface is more complicated than our examples of genus 2, and the construction will take some time. Moreover, since we cannot construct some famous surface without using some knowledge about it, we do not even have a well defined problem yet. One could start with the 24 tiles of the platonic tessellation by 120°-heptagons mentioned in the introduction. We found it interesting that Klein's surface is already determined by much less than its full symmetry, and by asking less we are rather naturally led to a fourteengon as a fundamental domain together with the correct identifications. The heptagons then fit into this fundamental domain in a way that can be described easily and platonicity follows with short arguments.

In analogy to the first genus 2 example we will look for a genus 3 surface tessellated by (rather big) $\pi/7$-triangles such that reflections in the edges extend to antiautomorphisms of the surface. There are only two such Riemann surfaces and both have a cyclic group of order 7 as automorphisms. But only one of the two has the 120° rotations around triangle centers as automorphisms. We construct one function by exploiting the order 7 rotation group and we find a second function with the Riemann mapping theorem. For both surfaces we derive an algebraic equation. For Klein's identification pattern we prove platonicity and finally complete the picture by a pairs of pants decomposition in terms of which all the remaining symmetries, in particular the symmetry subgroups, have simple descriptions.

4.1. Consequences of Euler's formula and of platonicity. First, for a given tessellation by $2\pi/3$-heptagons we obtain from Euler's formula the numbers F of faces, $V = \frac{7}{3} \cdot F$ of vertices and $E = \frac{7}{2} \cdot F$ of edges:

$$\chi(M^2) = -4 = F \cdot (1 - \tfrac{7}{2} + \tfrac{7}{3}) \quad \Longrightarrow \quad F = 24,\ V = 56,\ E = 84\,.$$

In the dual tessellation by $\frac{2\pi}{7}$-triangles the numbers F and V are interchanged. These numbers are too large to easily talk about individual tiles. By contrast, a tessellation by big $\pi/7$-triangles (of area $4\pi/7$ each) needs $F = 14$ of them to have the required total area 8π for a hyperbolic genus 3 surface; such a tessellation has $E = 21$ edges and $V = 3$ vertices.

Next consider a cyclic rotation group of prime order p on a surface of genus 3 with f fixed points. The Euler number for the quotient surface is given by

$$\frac{1}{p}\big((V - f) - E + F\big) + f \ = \ \frac{1}{p}(-4 - f) + f \in \{-2, 0, 2\},$$

or

$$f \in \left\{-2 + \frac{2}{p-1},\ \frac{4}{p-1},\ 2 + \frac{6}{p-1}\right\} \cap \mathbb{Z}.$$

Therefore $p = 7$ is the maximal prime order, $f = 3$ in that case and the quotient is a sphere. A genus 3 surface with an order 7 cyclic group of automorphisms therefore has a natural quotient map to the sphere. To view this map as a specific meromorphic function we identify the quotient sphere with $\mathbb{C} \cup \{\infty\}$ by sending the three fixed points to $0, 1, \infty$.

Furthermore, an involution ($p = 2$) must have $f = 0$, 4, or 8 fixed points. To discuss these possibilities further, note that an involution of a platonic tessellation by $2\pi/3$-heptagons cannot have its fixed points at vertices or centers of faces of the tessellation. Thus the fixed points are at edge midpoints. In such a case f must divide the number E of edges, therefore $f = 8$ cannot occur for an involution of our heptagon tessellation with E=84 edges — which shows in particular that the quotient is never a sphere, which is to say, Klein's surface is *not* hyperelliptic. We will see later that *all* involutions give quotient tori, $f = 0$ does not occur.

Platonicity further implies that we have a rotation group of order $p = 3$ around each of the heptagon vertices; we just computed its number f of fixed points:

$$f\big|_{p=3} \in \{2, 5\}.$$

Here $f = 5$ is excluded because it does not divide $V = 56$. So we have $f = 2$ and the quotient is a torus.

4.2. A fundamental domain from big triangles. Because of the desired cyclic symmetry group of order 7 we arrange the 14 big triangles around one center vertex to form a $2\pi/7$-fourteengon (Figure 4) and we see that all the odd and all the even vertices have to be identified to give a smooth hyperbolic surface. This leaves three possibilities: identify edge 1 to edge 4, 6 or 8. The last case has the 180°-rotation around the center as an involution with $f = 8$ fixed points (namely the center and the pairwise identified midpoints of fourteengon edges); that is, the quotient is a sphere. So this example is a hyperelliptic surface. As in the $\pi/5$-case we have found two quotient functions and their divisors give the equation

$$w^7 = z(z-1).$$

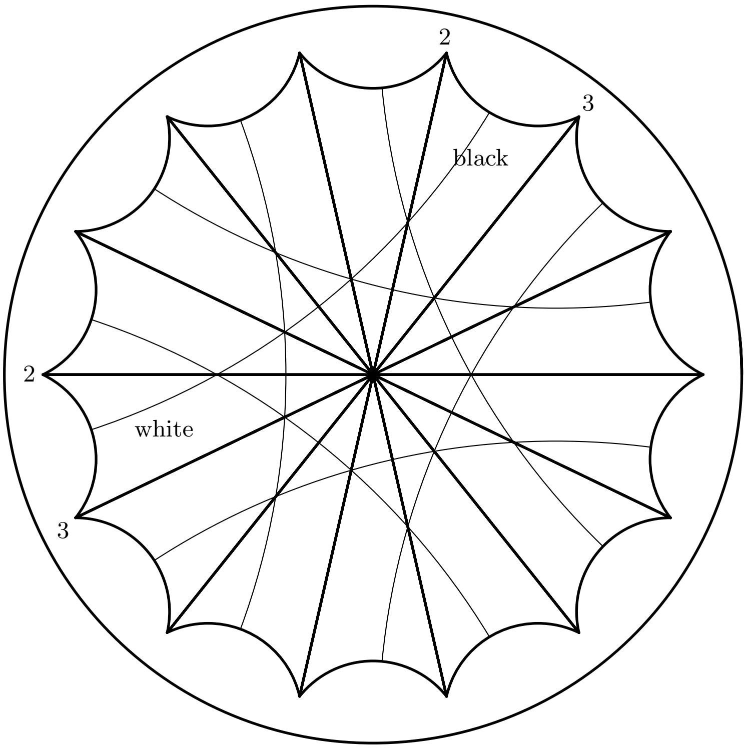


Figure 4. Hyperbolic fourteengon made from equilateral $\frac{\pi}{7}$-triangles, with translation axes.

Identification of edge 1 to 4 leads to the same hyperelliptic surface and therefore leaves the identification of edge 1 to edge 6 as the only candidate for some platonic surface, which we will prove to be Klein's surface. Note that this identification of the fourteengon edges is the hyperbolic description of the surface given in Klein's work; see the lithographic plate in [Klein 1879]. We concentrate on this case now and reveal further symmetries.

If one wants to check whether some expected symmetry is compatible with the identifications then the just given rule has the disadvantage that, for using it, one needs a rather large piece of the tessellation in the hyperbolic plane. We begin with a more convenient reformulation. Color the fourteen big triangles alternatingly black and white. Each black edge (of the fourteengon) is identified with the white edge that is counterclockwise 5 steps ahead (or the white edges with the black ones 9 steps ahead). We call the fourteengon center vertex 1, the left endpoint of a black edge vertex 2 and its right endpoint vertex 3. The identification rule can be restated as follows: Under the identification translation of a black edge to a white one, vertex 2 (as seen from vertex 1) is rotated by $2 \cdot 2\pi/7$ around the center and the triangle adjacent to this black edge is, at vertex 2, rotated by $1 \cdot 2\pi/7$; similarly, vertex 3 is rotated by $3 \cdot 2\pi/7$ around the center and the same triangle adjacent to this black edge is rotated around vertex 3 by $-1 \cdot 2\pi/7$. This can be expressed in a simpler way if one observes

$$2 \cdot \{1,2,4\} = \{2,4,1\} \pmod 7, \qquad 4 \cdot \{1,2,4\} = \{4,1,2\} \pmod 7.$$

The identification rule now is: Rotation around vertex 1 by $1 \cdot 2\pi/7$ is rotation at vertex 2 by $4 \cdot 2\pi/7$ and at vertex 3 by $2 \cdot 2\pi/7$. The high symmetry of Klein's surface is apparent in the fact that this rule remains the same $\pmod 7$ if we cyclically permute the vertices. — We remark that our description of Klein's surface in terms of flat cone metrics on a thrice punctured sphere will start from here.

To apply the new rule we consider a tessellation of the hyperbolic plane by the black and white $\pi/7$-triangles. Mark the equivalence classes of triangles from 1 to 14 and the vertices from 1 to 3, and observe that the identification rule allows us to pick an arbitrary triangle from each equivalence class and still know how to identify. The 120° rotation around any triangle center cyclically permutes the (equivalence classes of) vertices, but we saw that the identification rule is not affected by this change. Similarly, reflection in a triangle edge interchanges the black and white triangles and thereby the cyclic orientation of their vertices, but again, this does not change the identification rule. These reflections generate the order 7 rotational symmetry and therefore pass to the quotient sphere. This means that we can again understand the quotient map (under this symmetry group) via a Riemann mapping problem: Map a black triangle to the upper half plane, normalize so that the vertices 3, 2, 1 go to 0, 1, ∞ and extend by reflection.

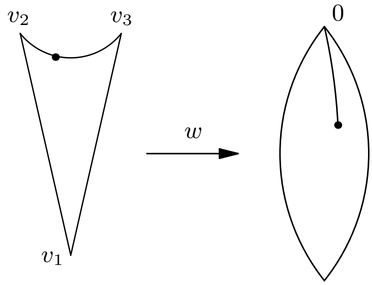


Figure 5. Equilateral hyperbolic $\frac{\pi}{7}$-triangle mapped to a spherical slit domain.

4.3. A second function and equations. We define a second function with the Riemann mapping theorem. Map one of the black triangles to a spherical domain that is bounded by two great circles from 0 to ∞ with angle $3 \cdot \pi/7$ at ∞ and has a great circle slit from 0 dividing the angle at 0 as $2:1$, counterclockwise the bigger angle first. (The length of the slit can be changed by scaling this map.)

This map can be extended analytically by reflection in the edges (around ∞) to cover the sphere three times. The slits in these three sheets are such that always in two sheets there are slits above each other, and these are not above a slit in the third sheet. This forced identification of the slits is compatible with the identifications of the edges of the fourteengon since the rotation angles $\{4, 1, 2\} \cdot 2\pi/7$ counterclockwise at the vertices of a black triangle are the same as the rotation angles $\{-3, 1, 2\} \cdot 2\pi/7$ at the vertices of the spherical domain. We compare the divisors of this function w and the above quotient function z and find that the functions w^7 and $z(z-1)^2$ are proportional. We can scale w to give us one of the known equations,

$$w^7 = z(z-1)^2 .$$

We do not know a group theoretic definition of the function w. Also, observe that the derivation did not use that this surface is platonic. Next we derive from this equation the even more famous quartic equation. It exhibits not only the order 7 symmetry (which was built in by construction) but gives another proof of the order 3 symmetry (independent of the above one). Consider this divisor table:

vertices	V_3	V_2	V_1
z	0^7	1^7	∞^7
w	0	0^2	∞^3
$v = w^2/(z-1)$	0^2	∞^3	0
$u = (z-1)/w^3$	∞^3	0	0^2
$\xi = v\,dz/z$	0	0^3	$\star$
$\omega = u\,dz/(1-z)$	0^3	$\star$	0
$\eta = u\,dz/(z(z-1))$	$\star$	0	0^3
$u\,z$	0^4	0	$\infty^5 r$

First, if we define

$$x := (1-z)/w^2 = -v^{-1}, \qquad y := -(1-z)/w^3 = +u$$

then the first equation implies the quartic equation

$$x^3\, y + y^3 + x = 0\,.$$

Of course, the substitution can be inverted: $w = -x/y, \quad z = 1 - x^3/y^2$. Secondly we see from the divisor table that the functions x, y are quotients of *holomorphic* 1-forms, namely

$$x = \xi/\omega, \quad y = \eta/\omega.$$

This observation gives an additional interpretation to the equation in its homogeneous form

$$\xi^3\eta + \eta^3\omega + \omega^3\xi \;=\; 0$$

as an equation between explicitly known holomorphic 1-forms. The projective embedding defined by this equation is called the canonical curve.

We see an order 3 symmetry as the permutation of the coordinates and an order 7 symmetry by multiplying ξ, η, ω with powers of a seventh root of unity, namely ζ^1, ζ^4, ζ^2.

The existence of the function u, v, w with single poles of order $3 < g+1$, and $u \cdot z$ of single pole of order 5, prove that V_1, V_2, V_3 are Weierstraß points with non-gap sequence $(3, 5, 6)$ and hence of weight 1. After platonicity is proved, we know that all the heptagon centers are such Weierstraß points. These are all since $g^3 - g = 24$ is the total weight.

4.4. The heptagon tessellation. We now add the heptagon tessellation to the previous picture. This will allow to prove platonicity with rather little effort. Notice that from now on the emphasis is on the involutions of the surface, they were not visible so far.

One $2\pi/3$-heptagon can be tessellated by fourteen $(2,3,7)$-triangles that fit together around its center. The big $\pi/7$-triangle has 24 times the area of one $(2,3,7)$-triangle. We now explain how to tessellate one (called "the first") of the black big triangles by 24 of the small $(2,3,7)$-triangles; compare Figure 6. Take half a heptagon (tessellated by seven of the small triangles) to the left of its diameter, with the vertex 1 at the upper end and half an edge from vertex 4 to the lower end of the diameter; now reflect the lowest $(2,3,7)$-triangle in the half edge to give us eight small triangles that already tessellate one third of the big triangle. (The lowest vertex will be the center of the fourteengon.) 120°-rotations around heptagon vertex 2 complete the desired tessellation of the big triangle. Now extend by reflections to Klein's tessellation of the hyperbolic plane by $(2,3,7)$-triangles and notice that these can be grouped either to a tessellation by heptagons or by big triangles, the vertices of the latter being centers of certain heptagons.

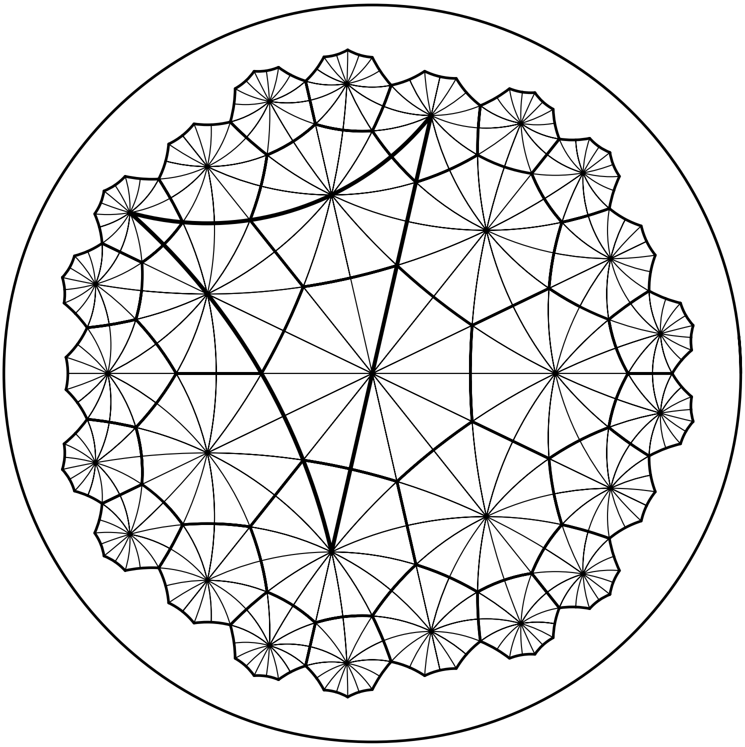


Figure 6. The tessellations by $\frac{\pi}{7}$-triangles and $\frac{\pi}{3}$-heptagons fit together.

Next we recover the above identification of the fourteengon edges as seen with the heptagons. The outer edge of the just tessellated big triangle joins counterclockwise fourteengon vertices, which we number 1 and 2. Connect midpoints of adjacent edges of the heptagon around the center of the fourteengon and extend these geodesics until they hit the fourteengon boundary. Notice that they are precisely eight such segments long. In other words, the eight segment translations along these geodesics give Klein's identification of the fourteengon edges! Notice also that the edges of the black triangles are indeed identified 5 steps forward (9 steps for the white ones). It is justified to quote Klein's lithographic plate again [Klein 1879].

We have now tessellated the above Riemann surface by 24 regular heptagons. Each vertex of the big triangle is the center of one heptagon and around each of these is a ring of seven heptagons. The identification translations are compositions of involutions (in the hyperbolic plane) around midpoints of heptagon edges that are four segments along a zigzag (called the Petrie polygon) apart — another indication that we have a platonic surface (Figure 7).

What remains to be checked? We know that the identification translations generate the deck group of a Riemann surface that also has the described order 3

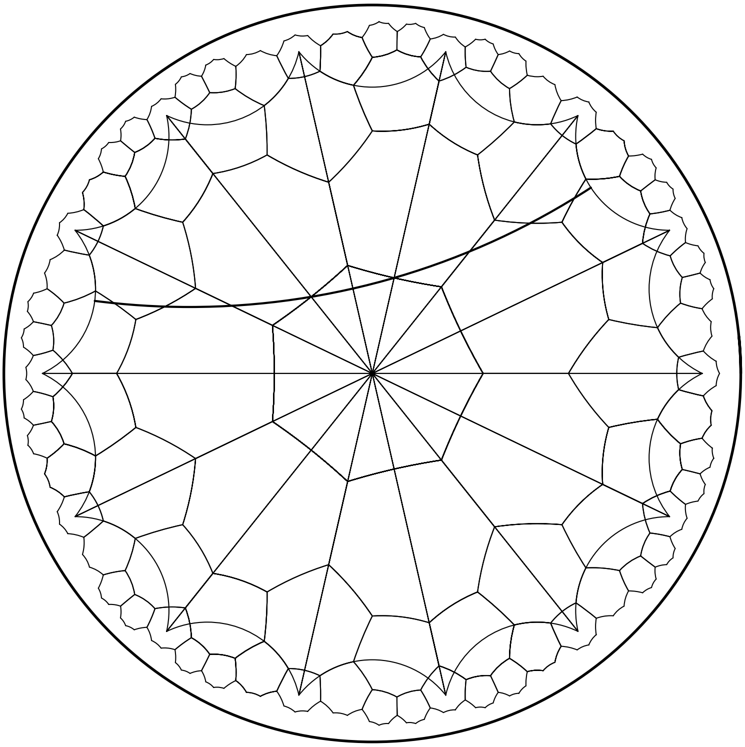


Figure 7. An eight-step geodesic crosses heptagon edge midpoints.

and order 7 rotational symmetries and the reflections in big triangle edges. Our Riemann surface will be platonic if *all* the eight segment geodesics that connect midpoints of heptagon edges in the hyperbolic plane connect equivalent points under the deck group, a new condition only for those that were not used to *define* the identifications. It is sufficient to check this for all the eight segment geodesics that meet the fourteengon fundamental domain. (If one discusses only candidates for identification generators and does not know a fundamental domain then at this point much more work is required.) Modulo reflections in symmetry lines through the center there are only four different eight segment geodesics that meet the fundamental domain. With the 120°-rotations these can be rotated into ones that were used to define the identifications! Now the hyperbolic description is complete enough to see platonicity, because the 180°-rotations around midpoints of heptagon edges in the hyperbolic plane always send equivalent points to equivalent ones.

How about other closed geodesics? If one connects the midpoints of second nearest heptagon edges and extends by applying 180°-rotations around the endpoints then these geodesics close after *six* such steps. Similarly, if one connects the midpoints of third nearest edges then these close again after eight steps.

Finally, also the symmetry lines close up: Let M be the midpoint of an edge, extend the edge across the neighboring heptagon, cross another heptagon, pass along another edge and cross a third heptagon to the midpoint M' of the opposite edge; M and M' are antipodal points of a closed geodesic (which is fixed under a reflection symmetry of the surface). All these geodesics are longer than the eight step ones used above and we did not see a fundamental domain that shows that one can take translations along them as generators of the deck group.

Finally, the heptagons also provide the connection with number theory: Puncture the surface in *all* the heptagon centers and choose a new complete hyperbolic metric by tessellating each punctured heptagon by seven $(3, 3, \infty)$-triangles. In the upper half plane model one such triangle is the well known fundamental domain for $SL_2(\mathbb{Z})$ and seven of them around the cusp at ∞ give the translation by 7 as one of the identification elements. This already connects Klein's surface with the congruence subgroup mod 7. In fact, $\Gamma(7)$ is the normal subgroup of the triangle group $SL_2(\mathbb{Z})$ (see Section 2) that is the deck group of this representation of Klein's surface. It is in this form that the surface first appears in [Klein 1879]; see [Rauch and Lewittes 1970; Gray 1982].

5. Oblique Pants and Isometry Subgroups

Pairs of pants decompositions are frequently used tools in the hyperbolic geometry of Riemann surfaces. One pair of pants is a Riemann surface of genus 0 bounded by three simply closed geodesics; it is further cut by shortest connections between the closed geodesics into two congruent right-angled hexagons. One builds Riemann surfaces by identifying pants along geodesics of the same length; the Fenchel–Nielsen coordinates are the lengths of these closed geodesics plus twist parameters since one can rotate the two pants against each other before the identification. If the hexagon vertices of neighboring pants coincide, the twist is $0°$ (or $180°$). Riemann surfaces have so many different pairs of pants decompositions that we need to say what we want to achieve for Klein's surface. The main motivation is to quickly understand symmetry subgroups that contain (many) involutions. The big triangle tessellation is not preserved under any involution and the heptagon tessellation has too many pieces. We find pairs of pants that are bounded by eight segment geodesics (the ones used in the previous section) in such a way that the twelve common vertices of the eight pant hexagons are fixed points of involutions. We will describe all types of symmetry subgroups with orders prime to seven in terms of this one pant decomposition. Our pant hexagons are not right angled but they have zero twist parameters. We have only found right angled pants with nonzero twists, therefore the oblique hexagons seem rather natural — to give a simpler example: one can subdivide parallelogram tori into rectangles, but only with a "twist"; that is, certain vertices lie on edges of other rectangles.

Our first goal is to develop a feeling for the shape of the surface — "versinnlichen" in Klein's words. Therefore we begin by giving a 1-parameter family of genus 3 surfaces, embedded in $\mathbb{R}^3$ with tetrahedral symmetry and with the full permutation group S_4 contained in the automorphism group. In this case one can visualize the rectangular quotient tori; this may help to appreciate the obliqueness of Klein's surface, which has no rectangular quotient tori.

Take a tetrahedral tessellation of the unit sphere and take a tube around its edge graph such that the tube not only respects the tetrahedral symmetry in $\mathbb{R}^3$ but also the conformal inversion in the unit sphere. Cut the legs between vertices by symmetry planes. This gives congruent pants, each with a 120° symmetry. The sphere cuts the pants into right-angled hexagons. One can interchange any two pants while mapping the others to themselves with a conformal map of $\mathbb{R}^3$ as follows: Invert in the unit sphere and then reflect in the plane of any of the great circle arcs of the tetrahedral graph with which we started.

An order 3 rotation subgroup commutes with the reflections in symmetry planes that contain the rotation axis, modulo this rotation group. The reflections pass to one reflection of the quotient torus, which clearly has two fixed point components. This makes the quotient torus rectangular. The 180°-rotations also commute with certain reflectional symmetries; they also descend with two fixed point components to the quotient torus.

In a last step we can ignore the embedding in $\mathbb{R}^3$ and identify the pants, using the *same twist* along all six closed geodesics. This will keep S_4 as a subgroup of automorphisms, but the complex conjugation on the quotient tori is generally lost. One of these surfaces is Klein's, another one is the Fermat quartic, see Section 6. For these special surfaces the quotient tori have again reflectional symmetries, but these are more difficult to imagine.

5.1. Pants for Klein's surface. Now we describe pants for Klein's surface in terms of the heptagon tessellation; see Figure 8. Because of the previous description we look for pants with a 120° symmetry. Select P, Q as fixed points of an order 3 rotation group. P, Q are opposite vertices of any pair of heptagons with a common edge e. We call e a symmetry line "between" P and Q; the edges "through" P or Q are not symmetry lines of the pants. Apply the rotations around P, Q to our first pair of heptagons. We obtain the three heptagons adjacent to P and the three adjacent to Q. Together they have the correct area for one pair of pants, and they are identified to a pair of pants along the three symmetry edges "between" P and Q, but the three pant boundaries are not yet closed geodesics, they are zig-zag boundaries made of eight heptagon edges. Next, extend the three edges from P slightly beyond the neighboring vertices until they orthogonally meet three of the closed eight segment geodesics. Observe that these three geodesics are also met orthogonally by the extension of the three symmetry edges between the heptagons around P and the heptagons around Q. This means that these three eight step closed geodesics cut a pair of pants out of

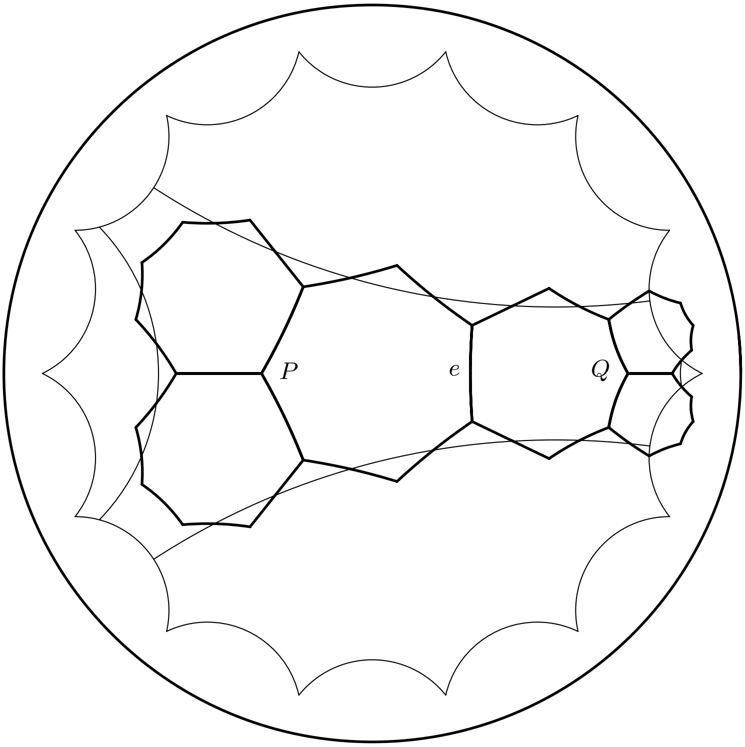


Figure 8. One pair of pants in the fourteengon fundamental domain.

the surface, which can be viewed as a smoothed version of the six heptagons. And the symmetry lines between P, Q cut these pants into two right-angled hexagons. (Precisely these pants have to be used in the initial description of an embedded surface. Since the hexagons of neighboring pants do not have common vertices we need a twist by one eighth of the total length of the boundary.)

Observe that reflection in the edges through P passes to a reflection of the quotient torus, quotient under the rotation group around P. This torus is made out of one pair of pants with two holes identified, the third hole is closed by one third of a pant, which is cut and identified along edges through P, Q. One can check that the fixed point set in the quotient has only one component so that the torus is rhombic. We find in Section 7 that its diagonals have a length ratio $\sqrt{7} : 1$.

Finally, we have to get the neighboring three pants, preferably with the help of an involution. Therefore we locate the *fixed points of one involution*: If one rotates around the midpoint M of any heptagon edge, then the two eight step geodesics through M are reversed so that their antipodal points $N, \overline{N}$ have also to be fixed points. Through both points there are again two eight step closed geodesics that get reversed; since the total number of fixed points is already

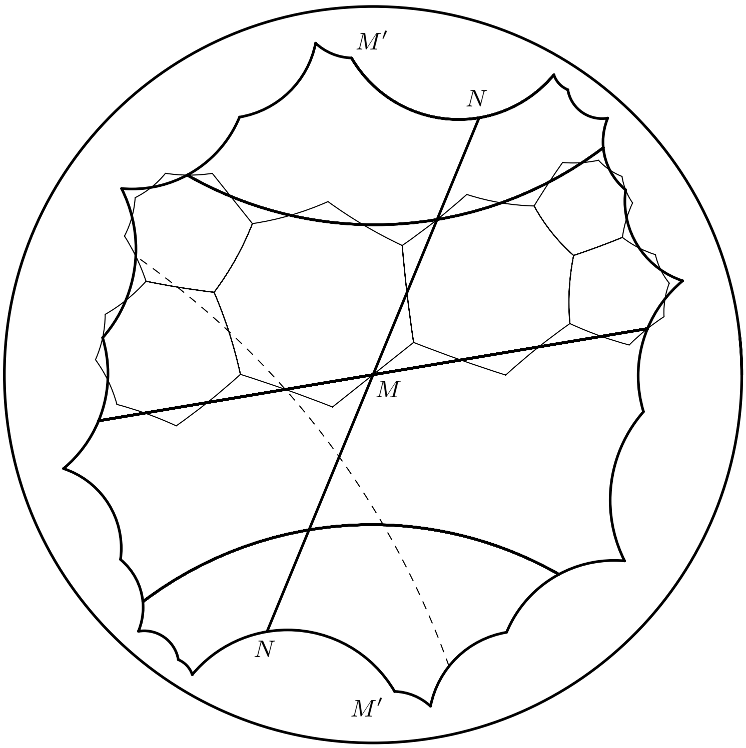


Figure 9. Eight big hexagons give four oblique pants and another fundamental domain.

known to be 4, these last two closed geodesics must meet in their common antipodal point M'.

We can use this intersection pattern of quadruples of eight step geodesics to get a pairs of pants decomposition without twist. First change the right angled hexagons to *oblique* ones: Instead of cutting the first pants into hexagons by the three symmetry edges between P and Q we use eight step geodesics through the midpoints of these edges, see Figure 9; at the first edge we have two choices, the other two are determined by the $120°$ symmetry. Observe that the sum of adjacent hexagon angles remains $= \pi$ and that the hexagon vertices are moved to involution centers! The edge lengths of the hexagons are now one quarter and one half, respectively, of our closed eight step geodesics. (Interpret the present description on the compact surface, but use a drawing in $\mathbb{H}^2$, Figure 9, to represent it.) The six vertices of the two hexagons of one pair of pants are three pairs of antipodal points on the three boundary geodesics — each of which consists of two long hexagon edges. Therefore each boundary geodesic can be rotated by $180°$ around the hexagon vertices on it and this moves the first pair of pants to three other pants, on the other side of each boundary geodesic. Note that the short edges that cut the first pair of pants into hexagons extend and also cut the neighboring pants into hexagons, i.e., these oblique pants fit together without twist and their hexagon vertices are involution centers. — We remark that the two conformal parameters of the family initially described in this section are in the present hyperbolic picture the ratio of adjacent hexagon angles and the ratio of adjacent edge lengths (recall the 120° symmetry of each hexagon).

So far the description emphasized an order 3 symmetry. We modify the description to emphasize a less obvious cyclic symmetry group of order 4; see Figure 9 above. Start with one of the closed geodesics, divided by involution centers M_i into four short hexagon edges (there are two possibilities for this subdivision, choose one). Through each of the four M_i choose the other closed geodesic, each divided by its antipodal point $\overline{M}_i$ into two long hexagon edges. (In the drawing in $\mathbb{H}^2$ each $\overline{M}_i$ is seen twice, each pair is connected by an eight step identification geodesic.) Through each of these four antipodes $\overline{M}_i$ we have again a unique other closed geodesic, but these are now pairwise the same ones — because we described above how four of them join the four fixed points of an involution. These last two geodesics therefore consist of the remaining short edges (sixteen in the $\mathbb{H}^2$-drawing) of our eight pant hexagons, so that we now have reached all the vertices. It remains to close the hexagons with long edges that fill up two more closed geodesics. We think of the hexagons as black and white, in a checkerboard fashion. Since there are $84 \cdot 2/8 = 21$ such closed geodesics we have $21 \cdot 2/3 = 14$ of these pant decompositions. Platonicity implies that the automorphism group is transitive on the set of 21 closed geodesics so that the isotropy group of each has order eight. We want to show that one such isotropy group leaves only one geodesic invariant. (Recall that the order 7 isotropy group of one heptagon has *three* invariant heptagons.) We can only propose proofs where the reader has to check how the eight step geodesics pass through a tessellation, by either pant hexagons or the earlier big triangles. Consider a tessellation by the big $\pi/7$-triangles that is kept invariant by a group of order 21, the order 7 rotations around the three common vertices and the fourteen order 3 rotations around the centers of the white big triangles (with the other fixed point of each rotation in the "opposite" black triangle). We claim that this group acts simply transitively on the 21 eight step closed geodesics. One can see this by following the geodesics that meet one of these triangles into the neighboring ones. Modulo its $\pm 120°$-rotations one white triangle is only met by three different kinds of eight step geodesics; already in one of the neighboring white triangles can one see that they are in fact all equivalent under this group.

5.2. Conjugacy classes and isometries. As a reward for this effort we can now describe all the isometries and also the subgroups of the automorphism group.

List of the conjugacy classes of the 168 orientation-preserving isometries. We have already characterized the isometries by sets that are left invariant; we only have to count that all 168 isometries have been found. Platonicity shows that all the isometries with the same characterization are in one conjugacy class.

Order 1: The class of the identity contains 1 element.

Order 2: The class of the involutions consists of 21 elements since each involution has four of the 84 edge midpoints as fixed points.

Order 3: The class of order three rotations has 56 elements since each of these rotations has two vertices as fixed points, and by interchanging the two fixed points with an involution one can conjugate one rotation and its inverse.

Order 7: There are two classes of order 7 rotations, each with 24 elements. Namely, with order 3 rotations can we cyclicly permute the three fixed points of one order 7 rotation and this conjugates the order 7 rotation with its second and fourth power; this gives two classes of three elements for each triple of fixed points, but each of the 24 heptagon centers can be mapped to every other one because of platonicity. The two classes of 24 elements are distinct, since only antiautomorphisms interchange black and white triangles.

Order 4: The class of order four translations of one eight step geodesic has 42 elements since each of the 21 closed geodesics has two such fixed point free translations and each translation is in the isotropy group of only one closed geodesic.

Altogether we have listed 168 isometries. So there are no isometries that we have not characterized, in particular no fixed point free involutions, and thus no genus 2 quotients of Klein's surface. The list also shows that the automorphism group is simple: Any normal subgroup has to consist of a union of full conjugacy classes, always including $\{id\}$; but its order has to divide 168, which is clearly impossible with the numbers from our list.

List of subgroups, assuming one fixed pairs of pants decomposition.

Order 2: Rotation around the midpoint of a short edge interchanges the adjacent black and white hexagons; every white hexagon has a black image. There are 21 of these subgroups.

Order 3: Cyclic rotation of one pair of pants into itself; cyclic permutation of the other pants. There are 28 of these subgroups.

Order 6: The two symmetries just given combine to the full isotropy group of *one* pair of pants. The decomposition into hexagons by short edges is not determined. There are 28 of these subgroups.

Order 4: From the construction of the pants we know the cyclic group generated by two step translations of a closed geodesic made of short edges. The uncolored tiling is preserved. There are 21 of these subgroups.

Order 4: The 180°-rotations around the twelve vertices of our pant hexagons form a Klein Four-group that preserves the colored tiling. There are 14 of these subgroups.

ORDER 8: Extend the cyclic translation subgroup of order 4 by the 180° rotation around the midpoint of one of the translated short edges. This is the isotropy group of an eight step closed geodesic. There are 21 of these subgroups.

ORDER 12: The full invariance group of the colored tessellation contains in addition to the above Klein group the order 3 rotations of each of the pants. There are 14 of these subgroups.

ORDER 24: All the above combine to the full invariance group of the uncolored tiling, abstractly this is the permutation group S_4. There are 14 of these subgroups.

ORDER 7: We know this as the invariance group of one heptagon. There are 8 of these subgroups.

ORDER 21: The invariance group of the tiling by 14 big triangles; no black and white triangles are interchanged; the isotropy of one triangle has order 3. There are 8 of these subgroups.

ORDER 14: Would contain an order 7 subgroup and an involution, hence at least 7 involutions and more order 7 rotations — too many.

ORDER 84: Would be a normal subgroup, which we excluded already.

For the remaining divisors of 168, namely 28, 42, 56, we have not found such a simple connection to the geometry. It is known that such subgroups do not occur, because an order 7 rotation and an involution generate the whole group.

6. Fermat Surfaces $x^k + y^k + z^k = 0$ Are Platonic

We add this section because, from the hyperbolic point of view, the Fermat quartic $x^4 + y^4 + z^4 = 0$ turns out to be surprisingly similar to Klein's surface. It has a platonic tessellation by twelve $2\pi/3$-octagons — one obtains the identification translations in the hyperbolic plane (which is tessellated by these octagons) if one joins two neighboring midpoints of edges and extends this geodesic to *six* such segments (see Figure 10 on page 35). Finally there is also a decomposition into congruent and 120°-symmetric pairs of pants that can be cut into oblique hexagons whose twelve common vertices are centers of involutions; this makes the Fermat quartic also a member of the 2-parameter family with at least S_4-symmetry, which we described in Section 5. Actually, *all* Fermat curves $x^k + y^k + z^k = 0$ can be described uniformly with their platonic tessellations. The hyperbolic picture is closer to this equation than in Klein's case, because the equation shows all the automorphisms immediately: One can independently multiply x and y by k-th roots of unity to obtain order k cyclic subgroups; cyclic permutation of the variables gives an order 3 rotation. In fact, any permutation of the variables gives an automorphism — including involutions, which were so hidden for Klein's surface. Also one can either derive from the equation the hyperbolic description or vice versa.

We start with a tessellated hyperbolic surface, point out obvious functions that have no common branch points and satisfy the Fermat equation: The rotations of order 3 and of order k point to a tessellation by π/k-triangles. To get their number we compute the Euler characteristic: The meromorphic function $f := x/z$ has k-fold zeros where the values of the function $g := y/z$ are k-th roots of -1; the k simple poles of both functions agree. The differential df has therefore $k \cdot (k-1)$ zeros and $2k$ poles, which gives $\chi = -k^2 + 3k$. Our expected triangle tessellation therefore has $F = 2k^2$ faces, $E = 3k^2$ edges and $V = 3k$ vertices and the dual tessellation consists of $3k$ regular $2k$-gons with angle $2\pi/3$. First we consider all tessellations in the hyperbolic plane. $2k$ of the triangles fit together around one vertex to form a regular $2k$-gon with angle $2\pi/k$. Into this we inscribe two regular k-gons with angle π/k by joining neighboring even-numbered and odd-numbered vertices, respectively. (Note that edges of these two polygons are symmetry lines of the triangle tessellation, the intersection of the two k-gons therefore is a regular $2k$-gon with angle $2\pi/3$.) We extend one of these inscribed k-gons (called red) to a tessellation of the hyperbolic plane and color its tiles in checkerboard fashion red and green; the other inscribed k-gon, called blue, we extend to a blue/yellow checkerboard tessellation. Note that the midpoints of the red and the blue k-gons agree; the vertices of the red/green ones are the midpoints of the yellow ones and vice versa, the vertices of the blue/yellow ones are the midpoints of the green ones. Now we define with the Riemann mapping theorem two functions on the hyperbolic plane, which we will recognize as the functions f, g above. Any checkerboard tessellation of the hyperbolic plane having an even number of regular polygons meeting at each vertex can be used in the same way: Map one tile to one hemisphere; we can keep its symmetry by first mapping a fundamental triangle (of the tiles symmetry group) to the corresponding sector of the hemisphere and then extend by reflection around the midpoint of the tile; finally extend by reflection in the edges of the tiles to the hyperbolic plane. We apply this by mapping a yellow and a green tile to the unit disc, normalizing so that the vertices go to k-th roots of -1. The functions, which we now call f and g, then have simple poles at the common centers of the red and blue polygons, respectively, and each has simple zeros at the other midpoints of its tiles, i.e., at the k-fold branch points of the other function whose branch values are k-th roots of -1. This gives the inhomogeneous Fermat equation

$$f^k + g^k + 1 = 0.$$

If we now identify points in the hyperbolic plane that are not separated by this pair of functions, then we are given a surface together with two tessellations by $2k$ regular k-gons; the vertices of both of them define a tessellation by $2k^2$ equilateral π/k-triangles. As a platonic tessellation the automorphism group would have to have order $6k^2$, but we already exhibited that many automorphisms of the Fermat equation — so this proves platonicity of the triangle tessellation and its dual, and then also of the tessellations by the k-gons. In particular this includes platonic tessellations with $\pi/5$- and $\pi/7$-triangles that we failed to obtain in the earlier attempts. Note also that $k = 3$ gives the triangle tessellation of the hexagonal torus.

We add some more details to the quartic, $k = 4$ (Figure 10). The k-gon tessellation consists of eight $\pi/4$-squares; they fit together around one vertex and give as fundamental domain a sixteengon with vertex angles alternately $\pi/4$ and $2\pi/4$ (and the vertices are identified to three points). There is only one possible edge identification pattern: If one wants the platonic symmetries around the center and notes the different angles at the vertices then the edge from a $\pi/4$-vertex clockwise to a $2\pi/4$-vertex can only be identified with the edges numbered 6 (translation axes in Figure 10) or 14 (clockwise). But the identifying translations are too short in the second case: The axis from edge 1 to edge 14 is two $\pi/4$-triangle edges long, which is only one half of a (vertex-)diameter of the sixteengon, a contradiction to platonicity. Now look at the dual of the triangle tessellation, by twelve $2\pi/3$-octagons (of which the figure shows nine) and note that we obtain the determined edge identification as composition of two involutions: Join two neighboring midpoints of edges of the central octagon and extend this (dithered) geodesic until it meets the boundary of the sixteengon fundamental domain. It is then six segments long; that is, the identification translation along this geodesic translates by a distance of six segments. Hence

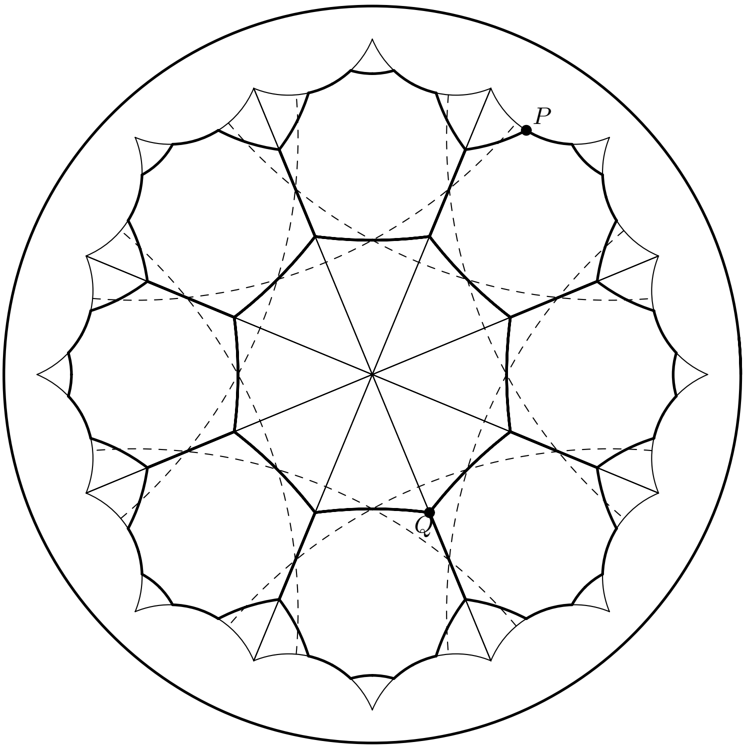


Figure 10. Fundamental domain for the Fermat quartic, with translation axes.

this translation can be written as a composition of two involutions whose centers are three segments apart, as claimed. This completes the hyperbolic description of the Fermat quartic with tiles and identifications.

We remark that the quotient by the 180°-rotation around the center is easy to see in both pictures: The fixed points are the four vertices of one tessellation by eight $\pi/4$-squares, the quotient therefore is the square torus tessellated by four squares; algebraically we have to identify points $(f,\, g) \sim (-f,\, g)$, i.e., we get the torus $x^2 + y^4 + 1 = 0$. Because of the order three symmetry we have three such quotient maps. An 180°-rotation around the midpoint of an octagon edge has also four fixed points so that the quotient is also a torus, a rectangular torus, because one reflection descends to the quotient with two fixed point components; for more information one has to compute. One may count that there are no other involutions, which proves that the Fermat quartic has no degree two projection to the sphere and therefore is not hyperelliptic; it also says that there are no fixed point free involutions; that is, our second genus 2 surface in Section 3.2, the one tessellated by the same $\pi/4$-triangles, is not a quotient of the Fermat quartic.

Finally we describe a pairs of pants decomposition such that the pant hexagons have involution centers as vertices. Number the vertices of the central $2\pi/3$-octagon of the fundamental domain. Choose vertex 6 as center of an order 3 rotation; the midpoint of the last edge (between vertex 8 and 1) is also midpoint between the *two* fixed points of this rotation; extend the octagon diameter from vertex 1 to 5 to a closed geodesic (a diameter of the fundamental domain) and rotate it by $\pm 2\pi/3$ around the chosen vertex 6; these three closed geodesics cut out of the surface a pair of pants tessellated by six half-octagons. Now we cut it into two hexagons: Join the midpoints of those two octagons, which have the mentioned edge from vertex 8 to 1 in common, across this edge; then also rotate this connection by $\pm 2\pi/3$ to obtain the desired hexagons. Finally obtain the neighboring pairs of pants by the involutions in the hexagon vertices — one does have to check that they do not overlap, but this is easy since the closed geodesics that we used to cut the pants into hexagons again traverse all four pairs of pants along short hexagon edges. The description of the symmetry subgroups is now very similar to the case of Klein's surface and will be omitted.

7. Cone Metrics and Maps to Tori

As we have seen above, certain quotients of Klein's surface are rhombic tori and we would like to know more about them. While we don't have any arguments using hyperbolic geometry to obtain this information, there is a surprisingly simple way using flat geometry. The idea is as follows: Suppose we have a holomorphic map from M^2 to some torus. Its exterior derivative will be a well defined holomorphic 1-form on M^2 with the special property that all its periods lie in a lattice in $\mathbb{C}$. Vice versa, the integral of such a 1-form will define a map to a torus whose lattice is spanned by the periods of the 1-form. There are two problems with this method:

It is rarely the case that one can write down holomorphic 1-forms for a Riemann surface. An exception are the hyperelliptic surfaces in their normal form $y^2 = P(x)$ where one can multiply the meromorphic form dx by rational functions in x and y to obtain a basis of holomorphic forms. But even if one can find holomorphic 1-forms then it is most unlikely that one can integrate them to compute their periods.

Flat geometry helps to overcome both problems simultaneously: Any holomorphic 1-form ω determines a flat metric on the surface that is singular in the zeroes of ω and that has trivial linear holonomy (parallel translation around any closed curve is the identity). This flat metric can be given by taking $|\omega|$ as its line element. Another way to describe it is as follows: Integrate ω to obtain a locally defined map from the surface to $\mathbb{C}$. Use this map to pull back the metric from $\mathbb{C}$ to the surface. A neighborhood around a zero of order k of ω is isometric to a euclidean cone with cone angle $2\pi(k+1)$, as can be seen in a local coordinate. And vice versa, specifying a flat *cone metric* without linear holonomy will always define a surface together with a holomorphic 1-form. The periods of this 1-form are just the translational part of the affine holonomy of the flat metric, which can be read off by developing the flat metric. Hence we have a method to construct Riemann surfaces with one holomorphic 1-form and full control over the periods. Usually one does not know whether two surfaces constructed by two different flat metrics coincide. The reason why we succeed with Klein's surface is that, surprisingly, we can apply the construction in three different ways so that we can produce three different 1-forms. This means that we have to show that three different cone metrics define the same conformal structure, which is difficult in general. But Klein's surface can be nicely described as a branched covering over the sphere with only three branch points, see Section 4.2, which allows to reduce this problem to the fact that there is only one conformal structure on the 3-punctured sphere.

For convenience, we introduce $[a, b, c]$-triangles, which are by definition euclidean triangles with angles

$$\frac{a\pi}{a+b+c}, \quad \frac{b\pi}{a+b+c}, \quad \frac{c\pi}{a+b+c}.$$

7.1. Again a definition of Klein's surface. Let $S = \mathbb{C}P^1 - \{P_1, P_2, P_3\}$ be a three punctured sphere. We construct a branched 7-fold covering over S that has branching order 7 at each P_i as follows: Choose another point P_0 in S and non-intersecting slits from P_0 to the punctures. Cut S along these slits, call the slit sphere S' and the edges at the slit from P_0 to P_i denote by a_i and a_i'. Now take 7 copies of S' and glue edge a_j in copy number i to edge a_j' in copy number $i + d_j \pmod 7$, where

$$d_1 = 1, \quad d_2 = 2, \quad d_3 = 4.$$

This defines a connected compact Riemann surface M^2 with a holomorphic branched covering map $\pi : M^2 \to S$. We call the branch points on M^2 also $P_1,\ P_2,\ P_3$. Viewing the sphere as the union of two triangles with vertices P_i, M^2 becomes then the union of fourteen triangles and Euler's formula gives $\chi(M^2) = -4$, $g = 3$. Equivalently, we could have used the Riemann–Hurwitz formula.

This description coincides with the one given in 4.2 — we have only switched from the identification of edges to slits for convenience. The reader can check again that the order 3 automorphism ϕ of the sphere that permutes the P_i lifts to M^2.

Observe that this description comes with a deck transformation of order 7.

7.2. Construction of holomorphic 1-forms. Now we want to construct holomorphic 1-forms on M^2. Consider a euclidean triangle with angles $\alpha_i \cdot \pi/7$ at the vertices P_i, for $i = 1, 2, 3$. Take the double to get a flat metric on the 3-punctured sphere, which also defines a conformal structure on the whole sphere. Because there is only one such structure, we can identify *any* doubled triangle with S and pull back the flat metric to M^2. In this metric a neighborhood of the branch points P_i on M^2 is isometric to a euclidean cone with cone angle $\alpha_i \cdot 2\pi$.

Remark that if we would take instead of the euclidean triangle a hyperbolic $2\pi/7$-triangle, this would give a hyperbolic metric without singular points — the same one that we know already.

After selecting a base point and a base direction on the universal cover $\hat{M}^2$ of $M^2 - \{P_1, P_2, P_3\}$, consider the developing map

$$\mathrm{dev} : \hat{M}^2 \to \mathbb{C}$$

of this flat metric. Let γ be a deck transformation of $\hat{M}^2$. Then $\mathrm{dev}(z)$ and $\mathrm{dev}(\gamma z)$ differ by an isometry of $\mathbb{C}$ and $\alpha(z)$ and $\alpha(\gamma z)$ with $\alpha = d\,\mathrm{dev}$ differ by a rotation. We want the holomorphic 1-form α to descend to M^2 and we therefore want all these rotations to be the identity. This is equivalent to having trivial linear holonomy of the flat cone metric on M^2. We call triangles such that the cone metric on M^2 above has this property *admissible.*

Because a triangle has a simpler geometry than a cone metric on M, we will do the holonomy computation on S and therefore need to be able to recognize closed curves on S that lift to closed curves on M^2:

Let c be a closed curve on S and $A_j = \#(c, a_j) =$ algebraic intersection number of c with the slit a_j. Let $\tilde{c}$ be any lift of c to M^2. Then $\tilde{c}$ is closed in M^2 if and only if $A_1 d_1 + A_2 d_2 + A_3 d_3 \equiv 0 \pmod 7$, because by crossing the slit a_j we change from copy i to copy $i + d_j$, the contributions from all crossed slits add up and we want to arrive in the same copy as we started.

To compute the linear holonomy of the curve c we modify it at every intersection with a slit as follows: Instead of crossing the slit a_j, we prefer to walk around the point P_j. The new curve will never cross a slit and therefore be ho-

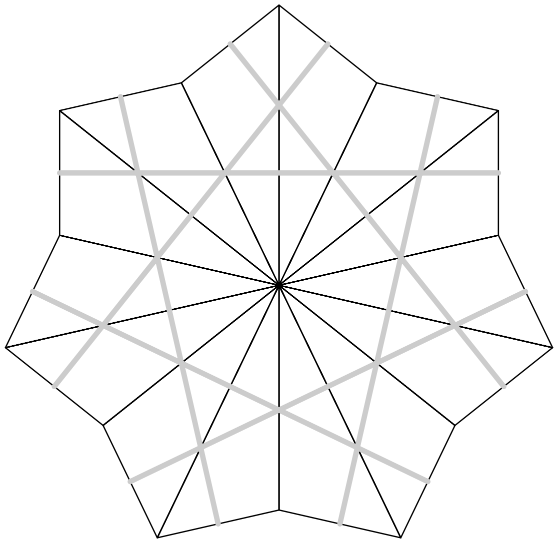


Figure 11. The flat fourteengon fundamental domain represents the surface together with a holomorphic 1-form.

motopically trivial, hence without linear holonomy. But each time we modified the curve c at the slit a_j, we changed the linear holonomy by a rotation by the cone angle $\alpha_j \cdot 2\pi/7$. This sums up to

$$\mathrm{hol}(c) = \text{rotation by } (A_1\alpha_1 + A_2\alpha_2 + A_3\alpha_3) \cdot \frac{2\pi}{7}\,.$$

Therefore, the linear holonomy of each closed curve in M^2 is trivial if and only if whenever $A_1 d_1 + A_2 d_2 + A_3 d_3 \equiv 0 \pmod 7$, then $A_1\alpha_1 + A_2\alpha_2 + A_3\alpha_3 \equiv 0 \pmod 7$. This is here the case for

$$(\alpha_i) = (1,2,4),\ (2,4,1),\ (4,1,2).$$

These are all the same triangles with differently labeled vertices. Corresponding to these three possibilities of choosing cone metrics we obtain *three different* holomorphic 1-forms ω_i on M^2.

All this is illustrated in Figure 11, which shows the fourteengon fundamental domain of Figure 4 where each $\pi/7$-triangle is replaced by an admissible euclidean triangle. That this metric gives rise to a holomorphic 1-form is instantly visible because the identifications of edges are achieved by euclidean parallel translations.

Observe that the scaling of an ω_i is well defined as soon as we have chosen a fixed triangle, but up to now there is no natural way and no necessity to do this.

As mentioned above, a cone angle $2\pi k$ of a cone metric causes a zero of order $k-1$ of the 1-form defined by the derivative of the developing map. So we obtain for the divisors of the ω_i on M^2 the expressions

$$(\omega_1) = P_2 + 3P_3\,,$$
$$(\omega_2) = P_1 + 3P_2\,,$$
$$(\omega_3) = P_3 + 3P_1\,.$$

This allows us to derive the equation for Klein's surface in a different way than in the hyperbolic discussion, because there the holomorphic 1-forms were only obtained after we had the first equation:

Set $f = \omega_1/\omega_3$ and $g = \omega_2/\omega_3$. We have

$$\begin{aligned}(f) &= -3P_1 + P_2 + 2P_3\,,\\ (g) &= -2P_1 + 3P_2 - P_3\,,\\ (g^3 f) &= -9P_1 + 10P_2 - P_3\,,\\ (f^3) &= -9P_1 + 3P_2 + 6P_3\,,\\ (fg^2) &= -7P_1 + 7P_2\,.\end{aligned}$$

Assuming that P_1 is mapped to ∞, P_2 to 0 and P_3 to -1 by π (see Section 7.1), so that $(\pi) = -7P_1 + 7P_2$ we see from the above table that (after scaling fg^2) fg^2 and π coincide. Therefore $fg^2 + 1$ has a zero of order 7 and $g^3 f + g$ a zero of order 6 in P_3. From the divisor table we see that there is only one pole of order 9 at P_1 which is completely compensated by the zeros at P_2 and P_3; hence $(g^3 f + g) = -9P_1 + 3P_2 + 6P_3$, and after a second normalization we have

$$f^3 + g^3 f + g \equiv 0 \quad \Longrightarrow \quad \left(\frac{\omega_1}{\omega_3}\right)^3 + \left(\frac{\omega_2}{\omega_3}\right)^3 \frac{\omega_1}{\omega_3} + \frac{\omega_2}{\omega_3} = 0,$$

so that the 1-forms themselves — suitably scaled — satisfy one equation for the Klein surface:

$$\omega_1\omega_2^3 + \omega_2\omega_3^3 + \omega_3\omega_1^3 = 0\ .$$

Note that the other equation can be written as

$$\pi(\pi - 1)^2 = f^7$$

by comparing divisors and scaling f.

7.3. Finding maps to tori. Now we want to find maps from M^2 to tori. First we determine the Jacobian. As explained above, we have to look for holomorphic 1-forms whose periods span a lattice in $\mathbb{C}$. Because M^2 is of genus 3, any holomorphic 1-form is a linear combination of the three forms ω_i above. We start by computing their periods. Because we have everything reduced to triangles, this is an exercise in euclidean geometry.

Consider an admissible triangle with angles $\alpha_i \cdot \pi/7$ in P_i and take the double. Choose the base point P_0 very close to P_1 and consider loops c_1, c_2 at P_0: c_1 is

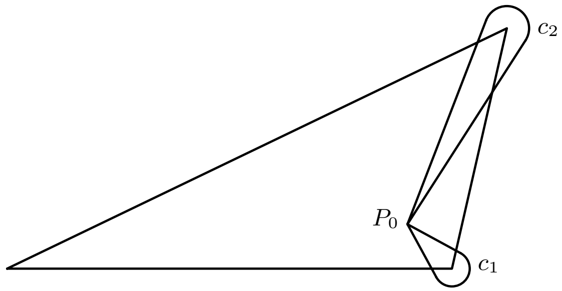


Figure 12. The integration paths.

a loop around P_1 and c_2 a loop around P_2, both with winding number 1. Every closed loop in S will be homotopic to a product of these two loops. Consider

$$\eta_k = c_1^k c_2 c_1^{5-k}, \quad \text{for } k = 1, \dots, 6.$$

These curves will have closed lifts to M^2 and it is easy to see that they furnish us with a homology base of M^2, for example as follows: Take the cone metric with $\alpha_1 = 1$ and repeatedly reflect the [1, 2, 4]-triangle to arrange all the fourteen copies around P_1. Since a lift of c_1 is a $2\pi/7$-arc around the center one can see that the $\eta_0, \dots, \eta_6$, are homotopic to the eight step closed geodesics that we used to identify edges of our fourteengon.

To compute their affine holonomy with respect to a cone metric on M^2, we can as well work on S. Recall how the development is constructed in this simple case: Follow the path starting at P_0 until it meets the boundary of the triangle (which is thought of as the upper hemisphere). Continue in the reflected triangle the portion of the path on the other hemisphere until it hits a triangle boundary again. Keep continuing until the endpoint of the path in S is reached and we have constructed the developped path in $\mathbb{C}$. This shows that $\operatorname{dev}(c_1)$ consists of a rotation by $\alpha_1 \cdot 2\pi/7$ and a translational part that can be made arbitrarily small since the holonomy is independent of how close we chose P_1 to P_0 . On the other hand, $\operatorname{dev}(c_2)$ consists (again up to an arbitrarily small error) of a translation by twice the height of the triangle with vertex P_1 followed by a rotation of angle $\alpha_2 \cdot 2\pi/7$. Since the last rotations do not change the endpoint of the developped path we obtain

$$\operatorname{dev}(\eta_k) = \zeta^{k\cdot\alpha_1} \cdot h_1 \quad \text{with} \quad \zeta = e^{2\pi \mathbf{i}/7}$$

and h_1 denotes the length of the height. Because we are still free to scale the triangles independently, we do this in a way that the periods look as simple as possible, namely we scale the height h_1 to length 1. So the triangles under consideration will have different size, but we obtain the periods as

$$\int_{\eta_k} \omega_1 = \zeta^k \,, \quad \int_{\eta_k} \omega_2 = \zeta^{2k} \,, \quad \int_{\eta_k} \omega_3 = \zeta^{4k} \,.$$

This gives a base for the lattice of the Jacobian of M^2.

As explained above, we now want to find linear combinations

$$\omega = a_1\omega_1 + a_2\omega_2 + a_3\omega_3$$

such that the periods of ω span a lattice in $\mathbb{C}$. Remark that if this is the case for some a_i, it will also be true for $\zeta^{d_i} a_i$. This has its geometric reason in the fact that the covering transformation of order 7 acts on the ω_i by multiplication with ζ^{d_i}. The corresponding maps to tori differ therefore only by an automorphism. We guess the first example of such a map to a torus:

The first map to a torus. Take $a_1 = a_2 = a_3 = 1$. Here $\omega = \omega_1 + \omega_2 + \omega_3$ and the periods $e_k := \int_{\eta_k} \omega$ lie in the lattice spanned by

$$v_1 = e_1 = e_2 = e_4 = \zeta^1 + \zeta^2 + \zeta^4 \quad \text{and} \quad v_2 = e_3 = e_5 = e_6 = \zeta^3 + \zeta^5 + \zeta^6\,.$$

Observe that

$$\left|\zeta^1 + \zeta^2 + \zeta^4\right|^2 = \left|\zeta^3 + \zeta^5 + \zeta^6\right|^2 = (\zeta^1 + \zeta^2 + \zeta^4)\cdot(\zeta^6 + \zeta^5 + \zeta^3) = 2\,;$$

hence

$$\left|\zeta^1 + \zeta^2 + \zeta^4 - \zeta^3 - \zeta^5 - \zeta^6\right|^2 = 7, \quad \zeta^1 + \zeta^2 + \zeta^4 = \frac{-1+\sqrt{-7}}{2}\,,$$

so that we obtain a map $\psi := \int \omega$ onto a rhombic torus $\mathbb{T}$ with edge length $\sqrt{2}$ and diagonal lengths $\sqrt{7}$ and 1. The lattice points are the ring of integers in the quadratic number field $\mathbb{Q}\big(\sqrt{-7}\big)$. This implies that the torus has complex multiplication: Multiplication by any integer in $\mathbb{Q}\big(\sqrt{-7}\big)$ maps the lattice into itself and therefore induces a covering of the torus over itself, in particular coverings of degree 2 and 7.

The standard basis for this lattice Γ is $\{1,\tau\}$, where τ is defined as $\frac{1}{2}\big(-1+\sqrt{-7}\big)$. The Weierstraß $\wp$-function for Γ is a degree 4 function for the index 2 sublattice $\tau\cdot\Gamma$ and $\wp(z/\tau)/\tau^2$ is the Weierstraß $\wp$-function for $\tau\cdot\Gamma$. Starting from these two functions one can derive the following equation for the torus, which is defined over $\mathbb{Q}$:

$$q'^2 = 7q^3 - 5q - 2\,.$$

Remember that we are hunting for the quotient tori of Klein's surface by the automorphism groups of order 2 and 3. Because we have scaled the triangles that define the ω_i to different size, it is unlikely that their sum will give us a 1-form invariant under the order 3 rotation. But we might have with ψ a degree 2 quotient map. To decide this, we compute the degree of ψ. With respect to the basis η_k of $H^1(M^2,\mathbb{Z})$ and the basis v_1, v_2 of $H^1(\mathbb{T},\mathbb{Z})$, we have the matrix representation

$$\psi_* = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

To switch to cohomology, we need the intersection matrix I for our homology basis and its inverse, which represents the cup product with respect to the dual basis. The first can be read off from Figure 4:

$$I = \begin{pmatrix} 0 & -1 & -1 & 0 & 0 & 1 \\ 1 & 0 & -1 & -1 & 0 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 0 & +1 & 1 & 0 & -1 & -1 \\ 0 & 0 & +1 & +1 & 0 & -1 \\ -1 & 0 & 0 & +1 & 1 & 0 \end{pmatrix}, \qquad I^{-1} = \begin{pmatrix} 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ -1 & -1 & 0 & -1 & 0 & -1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 1 & -1 & 0 & 0 \end{pmatrix}.$$

Denoting the dual basis of η_k by β_k and that of v_i by γ_i, we compute

$$\begin{aligned} \deg\psi &= \deg\psi \cdot \int_{\mathbb{T}} \gamma_1 \wedge \gamma_2 = \int_{M^2} \psi^*\gamma_1 \wedge \psi^*\gamma_2 \\ &= \int_{M^2} (\beta_1 + \beta_2 + \beta_4) \wedge (\beta_3 + \beta_5 + \beta_6) \\ &= 1 + 1 + 0 + 1 + 0 + 1 + 1 + 1 + 1 = 7\,. \end{aligned}$$

This is certainly a surprise, because we couldn't find any degree seven map to a torus in the hyperbolic setting. This means especially that ψ is *not* a quotient map.

The second map to a torus. As already mentioned, the above ω is not invariant under the triangle rotation automorphism ϕ of order 3, because we have normalized the ω_i using triangles of different size. By taking one fixed triangle size for all 1-forms, that is, by only permuting the labels of the vertices, we will obtain differently scaled 1-forms $\tilde\omega_i$, which now do have the invariance property

$$\phi^*\tilde\omega_i = \tilde\omega_{i+1}\,.$$

This means that

$$\int_{\eta_k} \tilde\omega_i = \zeta^{kd_i} \int_{\eta_0} \tilde\omega_i$$

with

$$\int_{\eta_0} \tilde\omega_i = \vec{h}_i\,,$$

where $\vec{h}_i \in \mathbb{C}$ is the height based at P_i in one fixed triangle $P_1P_2P_3$ with the angles $\beta_i := \alpha_i\pi/7$ at P_i, for $\alpha_i \in \{1, 2, 4\}$. Denote by h_i the norm of $\vec{h}_i$. Then compute

$$\vec{h}_2 = e^{-\mathbf{i}\beta_3}\frac{h_2}{h_1}\vec{h}_1, \qquad \frac{h_2}{h_1} = \frac{\sin\beta_1}{\sin\beta_2},$$

so that

$$\vec{h}_2 = e^{\mathbf{i}(\pi-\beta_3)}\frac{\sin\beta_1}{\sin\beta_2}\vec{h}_1\,, \quad \vec{h}_3 = e^{\mathbf{i}(\pi-\beta_1)}\frac{\sin\beta_2}{\sin\beta_3}\vec{h}_2\,.$$

Now introduce temporarily $\xi = e^{2\pi \mathbf{i}/14} = -\zeta^4$ and $\beta = \beta_1$. Write $\approx$ for equality up to a non-zero factor independent of k. Then

$$\begin{aligned}
\int_{\eta_k} \tilde{\omega}_1 + \tilde{\omega}_2 + \tilde{\omega}_3 = \zeta^k \vec{h}_1 + \zeta^{2k}\vec{h}_2 + \zeta^{4k}\vec{h}_3 \\
&= \left(\zeta^k - \zeta^{2k} e^{-\mathbf{i}\beta_3} \frac{\sin\beta_1}{\sin\beta_2} + \zeta^{4k} e^{-\mathbf{i}(\beta_1+\beta_3)} \frac{\sin\beta_1}{\sin\beta_3}\right) \vec{h}_1 \\
&= \left(\zeta^k - \zeta^{2k}\xi^{-4} \frac{\sin\beta}{\sin 2\beta} + \zeta^{4k}\xi^{-5}\frac{\sin\beta}{\sin 4\beta}\right)\vec{h}_1 \\
&\approx \frac{1}{\xi - \xi^{-1}}\xi^{2k} - \frac{1}{\xi^2 - \xi^{-2}}\xi^{4k-4} + \frac{1}{\xi^4 - \xi^{-4}}\xi^{8k-5} \\
&\approx \frac{\zeta^k}{\zeta - 1} + \frac{\zeta^{2k+2}}{\zeta^2 - 1} + \frac{\zeta^{4k-1}}{\zeta^4 - 1} \\
&\approx \frac{\zeta^k}{\zeta - 1} + \frac{\zeta^{2k}}{1 - \zeta^5} + \frac{\zeta^{4k}}{\zeta^5 - \zeta}\,.
\end{aligned}$$

Denote this last expression for the period over η_k by e_k. One easily computes

$$\begin{aligned}
e_2 &= 0\,, \\
e_0 &= -e_4 = e_1 - e_3\,, \\
e_5 &= -e_0 - e_1\,, \\
e_6 &= e_5 + 3e_0\,,
\end{aligned}$$

so that

$$\begin{aligned}
\tilde{v}_1 &:= -e_3 = 1 + \zeta^2 - \zeta^3 - \zeta^4\,, \\
\tilde{v}_2 &:= e_1 = -1 + \zeta^3 + \zeta^4 - \zeta^5
\end{aligned}$$

constitute a basis for the lattice spanned by all periods e_k. So this time we obtain a map $\tilde{\psi}$ to a torus as the quotient map $X \to X/(\phi)$. Using the above mapping degree argument, one finds indeed that $\deg\tilde{\psi} = 3$.

Remarkably, the quotients of the period vectors of the two tori agree:

$$v_1 \cdot \tilde{v}_2 = 2(\zeta^5 - \zeta^2) = v_2 \cdot \tilde{v}_1,$$

so that $\tilde{\psi}$ is a different map to the same torus $\mathbb{T}$.

The thrid map to a torus. Finally, we know two ways to find a degree 2 map to a torus. The first is to guess. This works as follows: Any holomorphic map $\psi : M^2 \to T^2$ will induce a complex linear map $\operatorname{Jac} M^2 \to \operatorname{Jac} T^2 = T^2$ and therefore a direct factor of $\operatorname{Jac} M^2$. After having found two such factors, there has to be a third so that the Jacobian of M^2 is up to a covering the complex product of three 1-dimensional tori. To find the third factor one just has to compute the kernel of the two linear maps that are the projections of $\operatorname{Jac} M^2$ to the tori already found and write down a projection onto this kernel. So the recipe is: Take the cross product of the 1-forms that define the maps to the two tori with respect to the basis ω_i.

For instance, we can take for the first torus the linear combination $\omega_1+\omega_2+\omega_3$ and for the second $\zeta\omega_1+\zeta^2\omega_2+\zeta^4\omega_3$, which is obtained from the first by applying an order 7 rotation. Hence also $(\zeta^4-\zeta^2)\omega_1+(\zeta-\zeta^4)\omega_2+(\zeta^2-\zeta)\omega_3$ integrates to a map to a torus, we compute the period integrals to

$$\begin{aligned} e_0 &= e_1 = 0\,, \\ -e_2 &= e_4 = \sqrt{-7}\,, \\ e_3 &= 3-\zeta^3-\zeta^5-\zeta^6 = \frac{7+\sqrt{-7}}{2}\,, \\ e_5 &= -3+\zeta+\zeta^2+\zeta^4 = -e_3+e_4\,. \end{aligned}$$

Taking $v_1 = e_4$ and $v_2 = e_3$ we get a basis for our familiar torus $\mathbb{T}$, and the computation of the mapping degree yields 2. Being a twofold covering, this map must be the quotient map of an involution.

The other way to find such a torus is analogous to the approach for the second torus: We just have to find a 1-form invariant under an involution. But while the order 3 rotations were apparent from the construction of the surface, this is not the case for the involutions, and we don't know a geometric method to derive the operation of an involution on the ω_i by euclidean or hyperbolic means. On the other hand, this operation already occurs in [Klein 1879] who used an algebraic-geometric description of his surface to obtain this map as

$$\begin{pmatrix} A \\ B \\ C \end{pmatrix} \longmapsto \frac{1}{\sqrt{-7}} \begin{pmatrix} -\zeta^2+\zeta^5 & \zeta^3-\zeta^4 & -\zeta\ +\zeta^6 \\ \zeta^3-\zeta^4 & -\zeta\ +\zeta^6 & -\zeta^2+\zeta^5 \\ -\zeta\ +\zeta^6 & -\zeta^2+\zeta^5 & \zeta^3-\zeta^4 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix},$$

where $A^3B+B^3C+C^3A=0$. This information can be used to obtain the above and other invariant 1-forms—one has only to be aware of the fact that one has to take the scaled 1-forms $\tilde{\omega}_i$, which satisfy $\tilde{\omega}_1\tilde{\omega}_2^3+\tilde{\omega}_2\tilde{\omega}_3^3+\tilde{\omega}_3\tilde{\omega}_1^3=0$.

7.4. Further computations for our examples. It is possible to compute the Jacobians of all the hyperbolic examples we have given in the preceding sections using cone metrics. This is quite straightforward. For instance, for the genus 2 surface constructed from hyperbolic $\pi/5$-triangles one uses as the conformal definition a 5-fold covering over the 3-punctured sphere analogous to step 1 above. Here one has to take $d_1 = 1$, $d_2 = 1$ and $d_3 = 3$. Using the same reasoning as in Section 7.2, one finds that admissible triangles for *this* covering are [1, 1, 3]- or [2, 2, 1]-triangles. This gives holomorphic 1-forms with divisors $2P_3$ and P_1+P_2, which can be used to derive an equation for the surface:

Introduce $w := \omega_2/\omega_1$ and denote the covering projection by z, normalized so that P_1, P_2, P_3 are mapped to 0, 1, ∞. After scaling w we obtain the equation from Section 3.1, $w^5 = z(z-1)$. The same computation as in 7.3 gives for the Jacobian of the surface the following basis matrix of the lattice

$$\begin{pmatrix} 1 & \zeta & \zeta^2 & \zeta^3 \\ 1 & \zeta^3 & \zeta & \zeta^4 \end{pmatrix} \quad \text{where} \ \ \zeta \ = e^{2\pi\mathbf{i}/5}.$$

The other genus 2-example is slightly more complicated. It is described as a double covering M^2 over a sphere punctured at octahedron vertices. So, instead of starting with the 3-punctured sphere as above we now have to start with a 6-punctured sphere. Here we can clearly see the limitations of our method: It is rarely the case that two different cone metrics on a 6-punctured sphere define the same conformal structure. However the octahedron itself is symmetric enough so that we can achieve this: Represent the conformal structure of the octahedron by the Riemann sphere with punctures at the images of the octahedron vertices under stereographic projection, that is at the fourth roots of unity, at 0 and at ∞. Then the map $z \mapsto z^4$ defines a branched covering over the 3-punctured sphere S, which we can handle. This means that any doubled triangle metric on S when lifted to the octahedron defines the same conformal structure. The metrics we obtain in this way can be described geometrically as follows: Instead of constructing the octahedron from equilateral triangles, it is allowed to construct it from isosceles triangles (bases along the equator). It is even allowed to take two different heights over the same base for the upper and the lower hemisphere. For instance, we can choose the triangles in such a way that the cone angles on the octahedron are π in ∞, 3π in 0 and also π at the roots of unity by taking four [2, 3, 3]-triangles and four [6, 1, 1]-triangles.

This cone metric on the octahedron is now admissible in the sense of Section 5.1: its lift to the double cover M^2 has no linear holonomy! This is an immediate consequence of these three facts:

– each branch point has order 2,
– each cone angle is an *odd* multiple of π,
– a closed curve on S has a closed lift to M^2 if and only if it crosses an even number of slits.

If we denote the branch points over 0 and ∞ by P_+ and P_-, respectively, we have found a holomorphic 1-form ω_1 with divisor $2P_+$. By interchanging the angles given to 0 and ∞, we obtain a 1-form ω_2 with divisor $2P_-$.

These two 1-forms are not sufficient to produce an equation for the surface. But this will be possible by using another *meromorphic* 1-form, also constructed using cone metrics: Represent the conformal structure of the octahedron just by the flat euclidean plane, where all cone points save ∞ have cone angle 2π and ∞ has -2π. The lift of this metric to M^2 (recall that all vertices are simple branch points) defines a meromorphic 1-form ω_3 with divisor $-3P_- + P_+ + P_1 + P_2 + P_3 + P_4$ where P_i denote the preimages of the roots of unity. Introduce the function $v = \omega_3/\omega_2$ and denote the covering projection from M^2 to the octahedron by z. Comparing divisors and scaling v now gives the equation

$$v^2 = z(z^4 - 1)$$

which is equivalent to the one in Section 3.2, put $w = v/((z+1)(z+\mathbf{i}))$.

Now we compute a basis for the lattice of the Jacobian of M^2. As a homology basis on M^2 we can take lifts of the 4 loops on the octahedron that start near 0 and go once around one root of unity. Because we have cone angle π at the roots of unity for both holomorphic 1-forms, these curves develop to straight segments of equal length, which we scale to 1. The directions can be easily obtained from the different cone angles at 0 and we get for the period matrix of ω_1, ω_2

$$\begin{pmatrix} 1 & \zeta & \zeta^2 & \zeta^3 \\ 1 & \zeta^3 & \zeta^6 & \zeta \end{pmatrix} = \begin{pmatrix} 1 & \frac{\sqrt{2}}{2}(1+\mathbf{i}) & \mathbf{i} & \frac{\sqrt{2}}{2}(-1+\mathbf{i}) \\ 1 & \frac{\sqrt{2}}{2}(-1+\mathbf{i}) & -\mathbf{i} & \frac{\sqrt{2}}{2}(1+\mathbf{i}) \end{pmatrix},$$

where $\zeta = e^{2\pi\mathbf{i}/8}$. From this it follows that $\omega_1+\omega_2$ and $\omega_1-\omega_2$ can be integrated to give maps to the isomorphic tori with lattices spanned by $\big(1, \sqrt{-2}/2\big)$ and $\big(1, 2\sqrt{-2}\big)$. They also have complex multiplication, as can be seen by folding a sheet of A4 paper.[1]

Similarly one can compute the Jacobians of all the Fermat surfaces. We carry this out for the quartic:

Here X will always be the four punctured sphere with punctures at the points $1, \mathbf{i}, -1, -\mathbf{i}$ or $0, 1, -1, \infty$. This is the only four punctured sphere for which we can sometimes describe *different* admissible cone metrics explicitly.

We will construct a branched covering M^2 of genus 3 over X very similar to the construction of Klein's surface, but this time using 4 slits instead of 3 and taking only a fourfold covering. We choose all the four numbers d_i that we need to specify the identifications to be 1. Using the Riemann–Hurwitz formula one can check that the so-defined surface has genus 3. Now one has to be careful to choose cone metrics on X, because we have to guarantee that different cone metrics live on the *same* 4-punctured sphere, namely X. This is done economically by representing X as a double cover over S such that $\mathbf{i}$, $-\mathbf{i}$ are mapped to ∞ and 1, -1 are mapped to 0, 1 without branching. Then *admissible* triangles on S in the sense that their lift to M^2 via X has no linear holonomy are given as [1, 5, 2]-, [5, 1, 2]- and $[2, 2, 4]$-triangles. These lift to three cone metrics on X with the following angles:

1	$\mathbf{i}$	-1	$-\mathbf{i}$
$\pi/2$	$\pi/2$	$5\pi/2$	$\pi/2$
$5\pi/2$	$\pi/2$	$\pi/2$	$\pi/2$
π	π	π	π

Counting the branching orders, this gives holomorphic 1-forms on M^2 with divisors $4P_2$, $4P_1$, and $P_1+P_2+P_3+P_4$. And they can be used to derive the equation $x^4+y^4+z^4=0$.

[1]The ISO series of paper sizes, A1, A2, ..., has the property that cutting in half an An sheet yields two A$(n+1)$ sheets similar to the original one.

For the computation of the Jacobian we want to use the curves c_1, c_2 defined in Section 7.3 and use the lifts of the curves

$$\eta_k = c_1^k c_2 c_1^{5-k} \quad \text{for } k = 0, \ldots, 5$$

to M^2 via X as a homology basis. Remark that these curves have closed lifts on M^2 because we have decided to start at P_1, which is an eightfold branch point of M^2 over 0 — starting at P_2 or P_3 would not produce closed curves. But having been careful gives after checking that the η_k form indeed a homology basis the following period matrix of the ω_i:

$$\begin{pmatrix} 1 & \zeta & \mathbf{i} & \zeta^3 & -1 & \zeta^5 \\ 1 & \zeta^5 & \mathbf{i} & \zeta^7 & -1 & \zeta \\ 1 & \mathbf{i} & -1 & -\mathbf{i} & 1 & \mathbf{i} \end{pmatrix}, \quad \zeta = e^{2\pi\mathbf{i}/8}.$$

So we see that ω_3, $\omega_1 + \omega_2$ and $\omega_1 - \omega_2$ integrate to maps to the square torus — which gives another proof that this M^2 does not cover the second genus 2-example: If this were the case, there would be a nontrivial map between their respective Jacobians by the universal property of Jacobians, inducing a nontrivial map from the square torus to the A4-torus (page 47, footnote). Such a map does not exist.

References

[Bujalance and Singerman 1985] E. Bujalance and D. Singerman, "The symmetry type of a Riemann surface", *Proc. London Math. Soc.* (3) **51**:3 (1985), 501–519.

[Burnside 1911] W. Burnside, *Theory of groups of finite order*, Cambridge U. Press, 1911.

[Coxeter and Moser 1957] H. S. M. Coxeter and W. O. J. Moser, *Generators and relations for discrete groups*, Springer, Berlin, 1957. Fourth edition, 1980.

[Gray 1982] J. Gray, "From the history of a simple group", *Math. Intelligencer* **4**:2 (1982), 59–67. Reprinted in this collection.

[Hurwitz 1893] A. Hurwitz, "Über algebraische Gebilde mit eindeutigen Transformationen in sich", *Math. Annalen* **41** (1893), 403–442.

[Klein 1879] F. Klein, "Ueber die Transformationen siebenter Ordnung der elliptischen Funktionen", *Math. Annalen* **14** (1879), 428–471. Reprinted as [Klein 1923, LXXXIV, pp. 90–136]. Translated in this collection.

[Klein 1923] F. Klein, *Gesammelte Mathematische Abhandlungen, 3: Elliptische Funktionen* etc., edited by R. Fricke et al., Springer, Berlin, 1923. Reprinted by Springer, 1973.

[Kulkarni 1982] R. S. Kulkarni, "Pseudofree actions and Hurwitz's $84(g-1)$ theorem", *Math. Annalen* **261**:2 (1982), 209–226.

[Lehner and Newman 1967] J. Lehner and M. Newman, "On Riemann surfaces with maximal automorphism groups", *Glasgow Math. J.* **8** (1967), 102–112.

[Macbeath 1965] A. M. Macbeath, "On a curve of genus 7", *Proc. London Math. Soc.* (3) **15** (1965), 527–542.

[Rauch and Lewittes 1970] H. E. Rauch and J. Lewittes, "The Riemann surface of Klein with 168 automorphisms", pp. 297–308 in *Problems in analysis* (Princeton, 1969), edited by R. C. Gunning, Princeton Univ. Press, Princeton, N.J., 1970.

[Rodríguez and González-Aguilera 1994] R. E. Rodríguez and V. González-Aguilera, "Fermat's quartic curve, Klein's curve and the tetrahedron", pp. 43–62 in *Extremal Riemann surfaces* (San Francisco, 1995), edited by J. R. Quine and P. Sarnak, Contemp. Math. **201**, Amer. Math. Soc., Providence, RI, 1997. Preprint, 1994.

[Schulte and Wills 1985] E. Schulte and J. M. Wills, "A polyhedral realization of Felix Klein's map $\{3,7\}_8$ on a Riemann surface of genus 3", *J. London Math. Soc.* (2) **32**:3 (1985), 539–547.

[Singerman 1995] D. Singerman, "Mirrors on Riemann surfaces", pp. 411–417 in *Second International Conference on Algebra* (Barnaul, 1991), edited by L. A. Bokut' et al., Contemp. Math. **184**, Amer. Math. Soc., Providence, RI, 1995.

[Streit 1996] M. Streit, "Homology, Belyĭ functions and canonical curves", *Manuscripta Math.* **90**:4 (1996), 489–509.

Hermann Karcher
Mathematisches Institut der Universität Bonn
Beringstrasse 6
53115 Bonn
Germany
unm416@ibm.rhrz.uni-bonn.de

Matthias Weber
Mathematisches Institut der Universität Bonn
Beringstrasse 6
53115 Bonn
Germany
weber@math.uni-bonn.de

The Eightfold Way
MSRI Publications
Volume **35**, 1998

The Klein Quartic in Number Theory

NOAM D. ELKIES

Abstract. We describe the Klein quartic $\mathcal{X}$ and highlight some of its remarkable properties that are of particular interest in number theory. These include extremal properties in characteristics 2, 3, and 7, the primes dividing the order of the automorphism group of $\mathcal{X}$; an explicit identification of $\mathcal{X}$ with the modular curve $X(7)$; and applications to the class number 1 problem and the case $n = 7$ of Fermat.

Introduction

Overview. In this expository paper we describe some of the remarkable properties of the *Klein quartic* that are of particular interest in number theory. The Klein quartic $\mathcal{X}$ is the unique curve of genus 3 over $\mathbb{C}$ with an automorphism group G of size 168, the maximum for its genus. Since G is central to the story, we begin with a detailed description of G and its representation on the three-dimensional space V in whose projectivization $\mathbb{P}(V) = \mathbb{P}^2$ the Klein quartic lives. The first section is devoted to this representation and its invariants, starting over $\mathbb{C}$ and then considering arithmetical questions of fields of definition and integral structures. There we also encounter a G-lattice that later occurs as both the period lattice and a Mordell–Weil lattice for $\mathcal{X}$. In the second section we introduce $\mathcal{X}$ and investigate it as a Riemann surface with automorphisms by G. In the third section we consider the arithmetic of $\mathcal{X}$: rational points, relations with the Fermat curve and Fermat's "Last Theorem" for exponent 7, and some extremal properties of the reduction of $\mathcal{X}$ modulo the primes $2, 3, 7$ dividing $\#G$. In the fourth and last section, we identify $\mathcal{X}$ explicitly with the modular curve $X(7)$, describe some quotients of $\mathcal{X}$ as classical modular curves, and report on Kenku's use of one of these quotients in a novel proof of the Stark–Heegner theorem on imaginary quadratic number fields of class number 1. We close that section with Klein's identification of $\pi_1(\mathcal{X})$ with an arithmetic congruence subgroup of $\mathrm{PSL}_2(\mathbb{R})$, and thus of $\mathcal{X}$ with what we now recognize as a Shimura curve.

Notations. We reserve the much-abused word "trivial" for the identity element of a group, the 1-element subgroup consisting solely of that element, or a group representation mapping each element to the identity.

Matrices will act from the left on column vectors.

We fix the seventh root of unity

$$\zeta := e^{2\pi i/7}, \tag{0.1}$$

and set

$$\alpha := \zeta + \zeta^2 + \zeta^4 = \frac{-1+\sqrt{-7}}{2}. \tag{0.2}$$

The seventh cyclotomic field and its real and quadratic imaginary subfields will be called

$$K := \mathbb{Q}(\zeta), \quad K_+ := \mathbb{Q}(\zeta+\zeta^{-1}), \quad k := \mathbb{Q}\big(\sqrt{-7}\big) = \mathbb{Q}(\alpha). \tag{0.3}$$

These are all cyclic Galois extensions of $\mathbb{Q}$. The nontrivial elements of $\mathrm{Gal}(K/\mathbb{Q})$ fixing k are the Galois automorphisms of order 3 mapping ζ to ζ^2, ζ^4; the nontrivial Galois automorphism preserving K_+ is complex conjugation $x \leftrightarrow \bar{x}$. As usual we write O_F for the ring of integers of a number field F; recall that O_K, O_{K_+}, O_k are respectively the polynomial rings $\mathbb{Z}[\zeta]$, $\mathbb{Z}[\zeta+\zeta^{-1}]$, $\mathbb{Z}[\alpha]$.

We use G throughout for the second-smallest noncyclic simple group

$$\mathrm{PSL}_2(\mathbb{F}_7) \cong \mathrm{SL}_3(\mathbb{F}_2)\ [\, = \mathrm{GL}_3(\mathbb{F}_2)\,] \tag{0.4}$$

of 168 elements.

Acknowledgements. Many thanks to Silvio Levy for soliciting this paper for the present MSRI volume and for his patience during repeated delays in the paper's completion.

I am grateful to Allan Adler, Benedict Gross, Barry Mazur, and J.-P. Serre for introducing me to many of the remarkable properties of the Klein quartic and for numerous enlightening conversations on various aspects of the geometry and arithmetic of $\mathcal{X}$ and of its automorphism group G. I also thank them, as well as Michael Bennett, Enrico Bombieri, Armand Brumer, Joe Harris, and Curt McMullen, for references to their and others' work and/or for clarifications of specific concepts and questions that arose in the process of putting this exposition together.

Hardly any of the results contained in this paper are original with me; some go back to Klein's work over a century ago, such as the explicit formulas for the representation of G and the determinantal expressions for its invariants [Klein 1879b], and the equations Kenku [1985] uses, referring to [Klein 1879a, § 7]. Much of the mathematical work of writing this paper lay in finding explicit equations that not only work locally to exhibit particular aspects of $(\mathcal{X}, G)$ but are also consistent between different parts of the exposition. The extensive symbolic computations needed to do this were greatly facilitated by the computer packages PARI and MACSYMA.

This work was made possible in part by funding from the National Science Foundation and the Packard Foundation.

1. The Group G and its Representation (V, ρ)

1.1. ***G* and its characters.** We reproduce from the ATLAS [Conway et al. 1985, p. 3] some information about G and its representations over $\mathbb{C}$. (That ATLAS page is also the source of facts concerning G cited without proof in the sequel.) The conjugacy classes c and character table of G are as follows:

c	$1A$	$2A$	$3A$	$4A$	$7A$	$7B$
$\#c$	1	21	56	42	24	24
χ_1	1	1	1	1	1	1
χ_3	3	-1	0	1	α	$\bar{\alpha}$
$\overline{\chi}_3$	3	-1	0	1	$\bar{\alpha}$	α
χ_6	6	2	0	0	-1	-1
χ_7	7	-1	1	-1	0	0
χ_8	8	0	-1	0	1	1

(1.1)

The outer automorphism group $\mathrm{Aut}(G)/G$ of G has order 2; an outer automorphism switches the conjugacy classes 7A,7B and the characters $\chi_3, \overline{\chi}_3$, and (necessarily) preserves the other conjugacy classes and characters. Having specified α in (0.2), we can distinguish χ_3 from $\overline{\chi}_3$ by labeling one of the conjugacy classes of 7-cycles as 7A; we do this by regarding G as $\mathrm{PSL}_2(\mathbb{F}_7)$ and selecting for 7A the conjugacy class of $\pm\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$. When we regard G as $\mathrm{PSL}_2(\mathbb{F}_7)$, the group $\mathrm{Aut}(G)$ is $\mathrm{PGL}_2(\mathbb{F}_7)$; if we use the $\mathrm{SL}_3(\mathbb{F}_2)$ description of G, we obtain an outer involution of G by mapping each 3×3 matrix to its inverse transpose.

Modulo the action of $\mathrm{Aut}(G)$ there are only two maximal subgroups in G (every other noncyclic simple group has at least three), of orders 21 and 24. These are the point stabilizers in the doubly transitive permutation representations of G on 8 and 7 letters respectively. These come respectively from the action of $G \cong \mathrm{PSL}_2(\mathbb{F}_7)$ on the projective line mod 7 and of $G \cong \mathrm{SL}_3(\mathbb{F}_2)$ on the projective plane mod 2. The 21-element subgroup is the normalizer of a 7-Sylow subgroup of G, and is the semidirect product of that subgroup (which is of course cyclic of order 7) with a group of order 3. Since all the 7-Sylows are conjugate under G, so are the 21-element subgroups, which extend to 42-element maximal subgroups of $\mathrm{Aut}(G)$ isomorphic to the group of permutations $x \mapsto ax+b$ of $\mathbb{F}_7$. The 24-element subgroup is the normalizer of a noncyclic subgroup of order 4 in G, and is the semidirect product of that subgroup with its automorphism group, isomorphic with the symmetric group S_3; thus the 24-element maximal subgroup is isomorphic with S_4. There are 14 such subgroups, in two orbits of size 7 under conjugation by G that are switched by an outer automorphism; thus these groups do not extend to 48-element subgroups of $\mathrm{Aut}(G)$.[1]

From these groups we readily obtain the irreducible representations of G with characters χ_6, χ_7, χ_8: the first two are the nontrivial parts of the 7- and 8-letter permutation representations of G, and the last is induced from a nontrivial one-dimensional character of the 21-element subgroup.

We now turn to χ_3 and $\overline{\chi}_3$. Let (V, ρ) and (V^*, ρ^*) be the representation with character χ_3 and its contragredient representation with character $\overline{\chi}_3$. Both V and V^* remain irreducible as representations of the 21- and 24-element subgroups; we use this to exhibit generators for $\rho(G)$ explicitly.

Fix an element g in the conjugacy class 7A. Then V decomposes as a direct sum of one-dimensional eigenspaces for $\rho(g)$ with eigenvalues ζ, ζ^2, ζ^4. The normalizer of $\langle g\rangle$ in G is generated by g and a 3-cycle h such that $h^{-1}gh = g^2$. Thus h cyclically permutes the three eigenspaces. The images of any eigenvector under $1, h, h^2$ therefore constitute a basis for V; relative to this basis, the matrices for $\rho(g), \rho(h)$ are simply

$$\rho(g) = \begin{pmatrix} \zeta^4 & 0 & 0 \\ 0 & \zeta^2 & 0 \\ 0 & 0 & \zeta \end{pmatrix}, \qquad \rho(h) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}. \tag{1.2}$$

In other words, the representation (V, ρ) restricted to the 21-element subgroup $\langle g, h\rangle$ of G is induced from a one-dimensional character of $\langle g\rangle$ sending g to ζ. Since this subgroup is maximal in G, we need only exhibit the image under ρ of some group element not generated by g, h. In his historic paper introducing (V, ρ) and his eponymous quartic curve, Klein [1879b, § 5] found that the involution

$$-\frac{1}{\sqrt{-7}} \begin{pmatrix} \zeta - \zeta^6 & \zeta^2 - \zeta^5 & \zeta^4 - \zeta^3 \\ \zeta^2 - \zeta^5 & \zeta^4 - \zeta^3 & \zeta - \zeta^6 \\ \zeta^4 - \zeta^3 & \zeta - \zeta^6 & \zeta^2 - \zeta^5 \end{pmatrix} \tag{1.3}$$

fills this bill. We thus refer to the image of G in $\mathrm{SL}_3(\mathbb{C})$ generated by the matrices (1.2,1.3) as the *Klein model* of (V, ρ).

The transformation (1.3) may seem outlandish, especially compared with (1.2), but we can explain it as follows. Except for the scaling factor $-1/\sqrt{-7}$, it is just the discrete Fourier transform on the space of odd functions $\mathbb{F}_7 \to \mathbb{C}$: identify such a function f with the vector $(f(1), f(2), f(4)) \in V$. It follows that this involution (1.3), as well as the transformations $\rho(g), \rho(h)$, are contained in Weil's group of unitary operators of the space of complex-valued functions on $\mathbb{F}_7$ [Weil

[1]Let H, H' be two subgroups of G isomorphic to S_4 in different orbits. Then H, H' are not conjugate in G, but are *almost conjugate* (a.k.a. "Gassmann equivalent" [Perlis 1977]): H, H' intersect each G-conjugacy class in subsets of equal size. Equivalently, the permutation representations of the action of G on the coset sets G/H, G/H' are isomorphic (in our case with character $\chi_6 \oplus \chi_1$). This has been used by Perlis to construct non-isomorphic number fields of degree 7 (the minimum) with the same zeta function [Perlis 1977] and, following [Sunada 1985], to exhibit isospectral planar domains [Gordon et al. 1992; Buser et al. 1994].

1964, § I]; they all commute with the parity involution $\iota : f(x) \leftrightarrow f(-x)$, and together generate the restriction to V of the commutator of ι in Weil's group. Starting with any odd prime p instead of 7, this would produce the $((p-1)/2)$-dimensional representation of $\mathrm{PSL}_2(\mathbb{F}_p)$ or of its double cover according as p is congruent to 3 or 1 mod 4; see also [Adler 1981, p. 116] for a concrete approach to the first case, of which G is the instance $p = 7$. If we take $g = \pm\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, $h = \pm\left(\begin{smallmatrix} 2 & 0 \\ 0 & 4 \end{smallmatrix}\right)$ in $\mathrm{PSL}_2(\mathbb{F}_7)$ then (1.3) is the image under ρ of the involution $s = \pm\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$.

The restriction of ρ to $S_4 \subset G$ is the group of orientation-preserving symmetries of the cube, that is, the group of signed 3×3 matrices of determinant 1. (The action on the four diagonals of the cube identifies this group with S_4; the 3-dimensional representation is the nontrivial part of the permutation representation of S_4 twisted by its sign character.) Unlike (V, ρ) and its restriction to the 21-element subgroup, this representation leaves a quadratic form invariant. We choose the subgroup isomorphic with S_4 generated by s, h, and $g^2 s g^{-2} = \pm\left(\begin{smallmatrix} 2 & 2 \\ 1 & -2 \end{smallmatrix}\right)$. Then the invariant quadric (which we shall need later) is a multiple of

$$X^2 + Y^2 + Z^2 + \bar{\alpha}(XY + XZ + YZ); \tag{1.4}$$

under the change of basis with matrix

$$\begin{pmatrix} 1 & 1+\zeta\alpha & \zeta^2+\zeta^6 \\ 1+\zeta\alpha & \zeta^2+\zeta^6 & 1 \\ \zeta^2+\zeta^6 & 1 & 1+\zeta\alpha \end{pmatrix} \tag{1.5}$$

we find that $s, h, g^2 s g^{-2}$ map to the signed permutation matrices

$$-\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \tag{1.6}$$

while g maps to

$$\frac{1}{2}\begin{pmatrix} -1 & 1 & \bar{\alpha} \\ \alpha & \alpha & 0 \\ -1 & 1 & -\bar{\alpha} \end{pmatrix}. \tag{1.7}$$

The matrices (1.6) and (1.7) generate an image of G in $\mathrm{SL}_3(\mathbb{C})$, which we shall call the S_4 *model* of (V, ρ).

We can also recover from (V, ρ) and (V^*, ρ^*) the irreducible representations of G of dimensions 6,7,8: the first is the symmetric square $\mathrm{Sym}^2(V)$; the second is $\mathrm{Sym}^3(V) \ominus V^*$; and the last is $(V \otimes V^*) \ominus \mathbf{1}$.

1.2. G-invariant polynomials in V. The action of G on V^* extends to an action on the ring

$$\mathbb{C}[V^*] = \bigoplus_{m=0}^{\infty} \mathrm{Sym}^m(V^*) \tag{1.8}$$

of polynomials on V. Klein determined over a century ago [1879b, § 6] the subring $\mathbb{C}[V^*]^G$ of polynomials invariant under this action: it is generated by three algebraically independent homogeneous polynomials of degrees 4, 6, 14, and a fourth polynomial of degree 21 whose square is a polynomial in the first three. It follows that the subring of polynomials invariant under the 336-element group $\pm G = \{\pm 1\} \times G$ is a polynomial ring generated by invariants of degrees 4, 6, 14. It is known [Shephard and Todd 1954] that a finite subgroup of $\mathrm{GL}_n(\mathbb{C})$ has a polynomial invariant ring if and only if it is a *complex reflection group*, that is, a group generated by its elements g such that $\mathbf{1}_n - g$ has rank 1. In our case the complex reflections in $\{\pm 1\} \times G$ are $-\rho(s)$ and its conjugates, of which there are 21 (the size of the conjugacy class $2A$). We next find explicit polynomials $\Phi_4, \Phi_6, \Phi_{14}, \Phi_{21}$ such that the invariant rings $\mathbb{C}[V^*]^{\pm G}$ and $\mathbb{C}[V^*]^G$ are generated by $\{\Phi_4, \Phi_6, \Phi_{14}\}$ and $\{\Phi_4, \Phi_6, \Phi_{14}, \Phi_{21}\}$ respectively, and determine Φ_{21}^2 as a polynomial in $\Phi_4, \Phi_6, \Phi_{14}$.

Letting $X, Y, Z \in V^*$ be the coordinate functions in the Klein model of (V, ρ), we can write the quartic invariant as

$$\Phi_4 := X^3Y + Y^3Z + Z^3X, \tag{1.9}$$

because even the action on $\mathrm{Sym}^4(V^*)$ of the 21-element subgroup of G generated by $(X, Y, Z) \mapsto (\zeta X, \zeta^4 Y, \zeta^2 Z)$ and cyclic permutations of X, Y, Z (see (1.2)) has only a one-dimensional invariant subspace, generated by Φ_4. The *Klein quartic* is the zero locus

$$\mathcal{X} := \{(X : Y : Z) \in \mathbb{P}(V) : \Phi_4(X, Y, Z) = 0\} \tag{1.10}$$

of Φ_4 in the projective plane $\mathbb{P}(V) = (V - \{\mathbf{0}\})/\mathbb{C}^*$. In the S_4 model the monomial matrices do not suffice to determine Φ_4 up to scaling, but starting from (1.9) we may use the change of basis (1.5) to find that Φ_4 is proportional to

$$X'^4 + Y'^4 + Z'^4 + 3\alpha(X'^2Y'^2 + X'^2Z'^2 + Y'^2Z'^2). \tag{1.11}$$

[We could also have determined the coefficient 3α by requiring invariance under the 7-cycle (1.7).] The formulas we exhibit[2] in the next three paragraphs for $\Phi_6, \Phi_{14}, \Phi_{21}$ in terms of Φ_4 can then be used to obtain those invariants as polynomials in the coordinates X', Y', Z' of the S_4 model, starting from (1.11).

Since Φ_4 is invariant under G, so is its Hessian determinant

$$H(\Phi_4) = \begin{vmatrix} \partial^2\Phi_4/\partial X^2 & \partial^2\Phi_4/\partial X\,\partial Y & \partial^2\Phi_4/\partial X\,\partial Z \\ \partial^2\Phi_4/\partial Y\,\partial X & \partial^2\Phi_4/\partial Y^2 & \partial^2\Phi_4/\partial Y\,\partial Z \\ \partial^2\Phi_4/\partial Z\partial X & \partial^2\Phi_4/\partial Z\,\partial Y & \partial^2\Phi_4/\partial Z^2 \end{vmatrix}, \tag{1.12}$$

[2]These determinantal formulas (1.13), (1.14), and (1.17) come straight from [Klein 1879b, §6]. Except for the coefficients 1/54, 1/9, 1/14, they can also be found in [Benson 1993, p. 101]; note that Benson's coordinates are related with ours by an odd permutation of the Klein coordinates X, Y, Z, and the 3×3 matrix for $\rho(s)$ in [Benson 1993] is missing the factor $1/\sqrt{-7}$ and has an incorrect $(3, 3)$ entry.

and we may take

$$\Phi_6 := -\frac{1}{54}H(\Phi_4) = XY^5 + YZ^5 + ZX^5 - 5X^2Y^2Z^2 \tag{1.13}$$

as the sextic invariant. These polynomials Φ_4, Φ_6 are f and $(-\nabla)$ in Klein's notation [1879b]. They are irreducible: each of Φ_4, Φ_6 can have at most 6 irreducible factors, permuted by G up to scaling, and since G has no proper subgroup of index ≤ 6 the factors must be themselves invariant; but the only invariant polynomials of degree < 4 are constant, so neither Φ_4 nor Φ_6 can admit a proper factorization.

The degree-14 invariant is not uniquely determined even up to scaling: one can also add any multiple of $\Phi_4^2\Phi_6$. But we will usually work mod Φ_4, so this additional ambiguity will disappear. A G-invariant polynomial of degree 14 not proportional to $\Phi_4^2\Phi_6$ can be obtained from either of the two conjugacy classes of subgroups $S_4 \subset G$: each of these contains seven subgroups, each of which has a unique invariant quadric (that is, an invariant line in $\mathrm{Sym}^2(V^*)$), and the product of these seven quadrics is a G-invariant polynomial of degree $7 \cdot 2 = 14$. We may choose for Φ_{14} any linear combination of this product and $\Phi_4^2\Phi_6$. Alternatively Φ_4 may be obtained as a differential determinant from Φ_4, Φ_6 by extending the Hessian determinant we used to obtain Φ_6 from Φ_4:

$$\Phi_{14} = \frac{1}{9}\begin{vmatrix} \partial^2\Phi_4/\partial X^2 & \partial^2\Phi_4/\partial X\,\partial Y & \partial^2\Phi_4/\partial X\,\partial Z & \partial\Phi_6/\partial X \\ \partial^2\Phi_4/\partial Y\,\partial X & \partial^2\Phi_4/\partial Y^2 & \partial^2\Phi_4/\partial Y\,\partial Z & \partial\Phi_6/\partial Y \\ \partial^2\Phi_4/\partial Z\,\partial X & \partial^2\Phi_4/\partial Z\,\partial Y & \partial^2\Phi_4/\partial Z^2 & \partial\Phi_6/\partial Z \\ \partial\Phi_6/\partial X & \partial\Phi_6/\partial Y & \partial\Phi_6/\partial Z & 0 \end{vmatrix}, \tag{1.14}$$

which in terms of the Klein coordinates for V is

$$\sum_{\mathrm{cyc}}(X^{14} - 34X^{11}Y^2Z - 250X^9YZ^4 + 375X^8Y^4Z^2 + 18X^7Y^7 - 126X^6Y^3Z^5) \tag{1.15}$$

(in which $\sum_{\mathrm{cyc}}$ means sum over the three cyclic permutations of X, Y, Z, so for instance $\Phi_4 = \sum_{\mathrm{cyc}} X^3Y$). All the invariant polynomials of degree 14 are irreducible except for $\Phi_4^2\Phi_6$ and the products of the two orbits of S_4-invariant quadrics. Multiplying the images of the quadric (1.4) under powers of $\rho(g)$ yields

$$\Phi_{14} + (69 + 7\alpha)\Phi_4^2\Phi_6, \tag{1.16}$$

so the reducible combinations of $\Phi_4^2\Phi_6$ and Φ_{14} are $\Phi_4^2\Phi_6$ itself, (1.16), and its conjugate $\Phi_{14} + (62 - 7\alpha)\Phi_4^2\Phi_6$.

Finally the invariant Φ_{21} may be described as the product of 21 linear forms: from the character table, each of the 21 involutions in G fixes a one-dimensional subspace of V^*, and we obtain Φ_{21} by multiplying generators of these subspaces. Alternatively Φ_{21} may be described as a multiple of the Jacobian determinant of $(\Phi_4, \Phi_6, \Phi_{14})$ with respect to (X, Y, Z). We choose the multiple

$$\Phi_{21} = \frac{\partial(\Phi_4, \Phi_6, \Phi_{14})}{14\,\partial(X, Y, Z)} = \frac{1}{14}\begin{vmatrix} \partial\Phi_4/\partial X & \partial\Phi_4/\partial Y & \partial\Phi_4/\partial Z \\ \partial\Phi_6/\partial X & \partial\Phi_6/\partial Y & \partial\Phi_6/\partial Z \\ \partial\Phi_{14}/\partial X & \partial\Phi_{14}/\partial Y & \partial\Phi_{14}/\partial Z \end{vmatrix}; \qquad (1.17)$$

the factor $1/14$ makes this an integral polynomial $X^{21} + Y^{21} + Z^{21} + \cdots$ in the Klein coordinates. Then Φ_{21}^2 is invariant under $\pm G$, and is thus a polynomial in $\Phi_4, \Phi_6, \Phi_{14}$. By comparing coefficients we find

$$\begin{aligned}\Phi_{21}^2 = \Phi_{14}^3 - 1728\Phi_6^7 + 1008\Phi_4\Phi_6^4\Phi_{14} - 32\Phi_4^2\Phi_6\Phi_{14}^2 + 19712\Phi_4^3\Phi_6^5 \\ - 1152\Phi_4^4\Phi_6^2\Phi_{14} + 11264\Phi_4^6\Phi_6^3 - 256\Phi_4^7\Phi_{14} + 12288\Phi_4^9\Phi_6. \qquad (1.18)\end{aligned}$$

Thus

$$\boxed{\Phi_{14}^3 - \Phi_{21}^2 \equiv 1728\Phi_6^7 \bmod \Phi_4.} \qquad (1.19)$$

The existence of a linear dependence mod Φ_4 between Φ_6^7, Φ_{14}^3, and Φ_{21}^2 could have been surmised from the degrees of these invariants; we shall see that it is closely related to the description of $\mathcal{X}$ as a G-cover of $\mathbb{CP}^1$ branched at only three points, with ramification indices $2, 3, 7$. (It is also the reason that this curve figures in the analysis of the Diophantine equation $Ax^2 + By^3 = Cz^7$ in [Darmon and Granville 1995].) The occurrence of the coefficient $1728 = 12^3$ in (1.19), reminiscent of the identity $E_2^3 - E_3^2 = 1728\Delta$ for modular forms on $\mathrm{PSL}_2(\mathbb{Z})$, suggests that $\mathcal{X}$ may be closely related with elliptic and modular curves; we shall see that this is in fact the case in the final section.

1.3. Arithmetic of (V, ρ): fields of definition. So far we have worked over $\mathbb{C}$. In fact all the representations of G except those of dimension 3 can be realized by homomorphisms of G to $\mathrm{GL}_d(\mathbb{Q})$; we say that these representations are *defined over* $\mathbb{Q}$. This is obvious for the trivial representation, and clear for the 6- and 7-dimensional ones from their relation with the 7- and 8-letter permutation representations of G. By comparing characters we see that the direct sum of the 7- and 8-dimensional representations is isomorphic with the exterior square of the 6-dimensional one, whence the 8-dimensional representation is also defined over $\mathbb{Q}$. We cannot hope for the 3-dimensional representations to be defined over $\mathbb{Q}$, because χ_3 takes irrational values $\alpha, \bar{\alpha}$ on the 7-cycles in G. We next investigate how close we can come to overcoming this difficulty.

The S_4 model shows that (V, ρ) can be defined over the quadratic extension k of $\mathbb{Q}$ generated by the values of χ_3. On the other hand, the Klein model of (V, ρ) uses matrices over the larger field K, but is defined over $\mathbb{Q}$ in the weaker sense that $\rho(G) \subset \mathrm{SL}_3(K)$ is stable under $\mathrm{Gal}(K/\mathbb{Q})$. Indeed the Galois conjugates of $\rho(g)$ are its powers, $\rho(h) \in \mathrm{SL}_3(\mathbb{Q})$ is fixed by $\mathrm{Gal}(K/\mathbb{Q})$, and the involution (1.3) is contained in $\mathrm{SL}_3(K_+)$ and taken by $\mathrm{Gal}(K_+/\mathbb{Q})$ to its conjugates by powers of h, so the group $\rho(G)$ generated by these three linear transformations is permuted by $\mathrm{Gal}(K/\mathbb{Q})$. The S_4 model cannot be defined over $\mathbb{Q}$ even in this weaker sense: if it were, complex conjugation would induce a nontrivial automorphism of G fixing $S_4 \subset G$ pointwise, but no such automorphism exists. This is why the invariants $\Phi_4, \Phi_6, \Phi_{14}, \Phi_{21}$ are polynomials over $\mathbb{Q}$ in the Klein model but not in the S_4 model. This still leaves open the possibility of finding a model in which $\rho(G)$ is both contained in $\mathrm{SL}_3(k)$ and stable under $\mathrm{Gal}(k/\mathbb{Q})$ by applying a suitable $\mathrm{GL}_3(K)$ or $\mathrm{GL}_3(k)$ change of basis to the Klein or S_4 model.

Indeed it turns out that such a model, giving in effect a faithful representation of $\mathrm{Aut}(G)$ into $\Gamma\mathrm{L}_3(k)$,[3] does exist, and is in fact unique up to isomorphism. This is because constructing such a model amounts to choosing an outer involution of G to map to the Galois involution of $k/\mathbb{Q}$, and there is just one conjugacy class of involutions in $\mathrm{Aut}(G) - G$. Under the identification of $\mathrm{Aut}(G)$ with $\mathrm{PGL}_2(\mathbb{F}_7)$, one such involution is $r = \pm\left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)$. The subgroup of G fixed by this involution is the copy of S_3 generated by h, s; thus only this subgroup will map to matrices in $\mathrm{GL}_3(\mathbb{Q})$. Allan Adler points out (in e-mail) a beautiful way to see the image of the 42-element subgroup $\langle g, h, r\rangle$ of $\mathrm{Aut}(G)$: regard K as a three-dimensional vector space over k; let g be multiplication by ζ; let h be generator of $\mathrm{Gal}(K/k)$ taking ζ to ζ^2; and let r be complex conjugation, acting k-antilinearly as it should. Since, as noted already, $\langle g, h\rangle$ acts irreducibly on V, this suffices to determine the representation. We choose the basis $(\zeta - \zeta^6, \zeta^2 - \zeta^5, \zeta^4 - \zeta^3)$ for K/k—note that this basis is orthogonal under the G-invariant Hermitian norm $\|\beta\| = \mathrm{Tr}_{K/k}(\beta\bar{\beta})$ on K. We find that this basis is related with the basis for the S_4 model by the change of basis with matrix

$$\begin{pmatrix} -\alpha & 1 & 2\alpha+3 \\ 2\alpha+3 & -\alpha & 1 \\ 1 & 2\alpha+3 & -\alpha \end{pmatrix}, \tag{1.20}$$

and that in this basis the matrices for $\rho(g), \rho(h), \rho(s)$ are

$$\frac{1}{\sqrt{-7}}\begin{pmatrix} -2 & \alpha & -1 \\ \alpha & -1 & 1-\alpha \\ -1 & 1-\alpha & -1-\alpha \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad \frac{1}{7}\begin{pmatrix} -3 & -6 & 2 \\ -6 & 2 & -3 \\ 2 & -3 & -6 \end{pmatrix}. \tag{1.21}$$

We call this the *rational S_3 model* of (V, ρ). Since it is weakly defined over $\mathbb{Q}$, its polynomial invariants have rational coefficients. For most purposes it is still more convenient to use the simpler invariants of the Klein model; for instance the quartic invariant Φ_4, which is the pretty trinomial (1.9) in the Klein model, becomes a multiple of

$$A^4+B^4+C^4+6(AB^3+BC^3+CA^3)-3(A^2B^2+B^2C^2+C^2A^2)+3ABC(A+B+C) \tag{1.22}$$

in our basis, and looks even worse with other coordinate choices. But it does have the advantage not only of minimal fields of definition but also of identifying

[3] By this is meant the semidirect product of $\mathrm{GL}_3(k)$ with $\mathrm{Gal}(k/\mathbb{Q})$, in analogy with the semilinear groups $\Gamma\mathrm{L}_n(\mathbb{F}_q)$ over finite fields properly containing $\mathbb{F}_p$.

G with linear groups over both $\mathbb{F}_2$ and $\mathbb{F}_7$ by reducing (V, ρ) modulo primes of O_K with those residue fields.

1.4. Arithmetic of (V, ρ): reduction mod p and the lattice L. Remarkably the representation (V, ρ) remains irreducible at every prime, and its reductions mod 2 and 7 reveal the identification of G with $\mathrm{SL}_3(\mathbb{F}_2)$ and $\mathrm{PSL}_2(\mathbb{F}_7)$ respectively. Before showing this we put it in context by briefly recalling what it means to reduce a representation mod p.

For this paragraph only, let G be any finite group, and (V, ρ) an irreducible representation of G defined over a number field F. Let $L \subset V$ be an O_F-lattice stable under G. (Such a lattice always exists; for instance we may choose any nonzero $v \in V$ and take for L the O_F-linear combinations $\sum_{g \in G} a_g \rho(g)(v)$.) For each prime ideal p of O_F, we then obtain a representation of G on the (O_F/p)-vector space L/pL. If this representation is irreducible then it does not depend on the choice of L, and we may unambiguously say that (V, ρ) is irreducible mod p and call L/pL its reduction mod p. This is the case for all but finitely many p, including all primes whose residual characteristic does not divide the order of G. But it may, and usually does, happen that there are some primes p, necessarily with $\#G \equiv 0 \bmod p$, such that L/pL is reducible, in which case that representation may depend on the G-stable lattice L (though the composition factors of L/pL depend only on (V, ρ) and p). For instance, if $F = \mathbb{Q}$ and G is the symmetric group S_n $(n > 3)$, and we take for (V, ρ) its usual $(n-1)$-dimensional representation, then it is reducible mod p if and only if p divides n. When p divides n, the representation L/pL depends on the choice of L. If we choose for L the root lattice

$$A_{n-1} = \left\{ (a_1, a_2, \ldots, a_n) \in \mathbb{Z}^n : \sum\nolimits_1^n a_i = 0 \right\},$$

the representation L/pL contains the 1-dimensional trivial representation generated by $(1, 1, \ldots, 1)$; if we choose instead the dual lattice A_{n-1}^* then L/pL has a G-invariant functional but no invariant proper subspace of positive dimension.

We return now to the case that G is the simple group of 168 elements and V is its 3-dimensional representation with character χ_3. We may choose either $F = K$ or $F = k$. In either case we may see without any computation that V is reducible mod p for each prime p of F. Indeed if V was reducible then G would have a nontrivial representation mod p of dimension 1 or 2; since G is simple and non-abelian, it would thus be a subgroup of $\mathrm{GL}_2(O_F/p)$. But the only non-abelian simple groups with an irreducible 2-dimensional representation over some field are the groups $\mathrm{SL}_2(\mathbb{F}_{2^r})$ for $r > 1$ (this follows from the classification of finite subgroups of SL_2 over an arbitrary field, see for instance [Suzuki 1982, Theorem 6.17]). But G is not such a group — it does not even have order $2^{3r} - 2^r$. This completes the proof that V is irreducible at each prime of F.

Thus (V, ρ) is one of the few known representations of finite groups in dimension greater than 1 that are "absolutely irreducible" in the sense of [Gross

1990], that is, are irreducible and remain so in every characteristic.[4] Since k has unique factorization, the main result (Prop. 5.4) of [Gross 1990] then shows that the lattice L is unique up to scaling. In the coordinates of the rational S_3 model L is proportional to the self-dual lattice

$$\left\{ \frac{1}{\sqrt{-7}}(x,y,z) : x,y,z \in O_k;\ y-2x, z-4x \in \sqrt{-7}\,O_K;\ x+2y+4z \in 7O_K \right\}. \tag{1.23}$$

In the coordinates of the S_4 model we may take L to be the O_k-lattice generated by the column vectors

$$(2,0,0), \quad (\alpha,\alpha,0) \quad (\bar{\alpha},1,1). \tag{1.24}$$

The group G can in turn be defined as the group of determinant-1 automorphisms of this lattice [Conway et al. 1985]. Likewise the only G-invariant lattices in V^* are of the form cL^* for nonzero c, where L^* is generated by

$$(2,0,0), \quad (\bar{\alpha},\bar{\alpha},0) \quad (\alpha,1,1); \tag{1.25}$$

this L^* may be identified with the dual lattice of L. (Of course L, L^* are isomorphic *qua* lattices because the representations V, V^* are identified by an automorphism of G.) We note two facts for future reference. First, that in our case it is enough to assume that L or L^* is a $\mathbb{Z}$-lattice stable under the action of G: we obtain the action of O_k automatically because $\rho(g)+\rho(g^2)+\rho(g^4)$ is multiplication by α on V and by $\bar{\alpha}$ on V^*. Second, that L is known to be the unique indecomposable positive-definite unimodular Hermitian O_k-lattice of rank 3 [Hoffmann 1991, Theorem 6.1].

We next consider the reductions of (V,ρ) in characteristics $2, 7$. We deal with characteristic 2 first. There are two primes $\wp_2, \bar{\wp}_2$ above 2 in O_k, interchanged by complex conjugation. We may take $\wp_2 = (\alpha)$, $\bar{\wp}_2 = (\bar{\alpha})$. Thus the reductions of the rational S_3 model for (V,ρ) modulo those primes are related by an outer automorphism of G. Using either prime, we obtain a nontrivial representation $G \to \mathrm{GL}_3(\mathbb{F}_2)$. Since G is simple, this map must be an isomorphism. That is, each invertible linear transformation of V mod $\wp_2$ or $\bar{\wp}_2$ comes from a unique element of G; equivalently, each automorphism of $L/\wp_2 L$ or $L/\bar{\wp}_2 L$ lifts to a unique determinant-1 isometry of L! Now Dickson proved that for each prime power q and every positive integer n the ring of invariants for the action of $\mathrm{GL}_n(\mathbb{F}_q)$ on its defining representation is polynomial, with generators of degrees $q^n - q^m$ for $m = 0, 1, \ldots, n-1$. (See the original paper [Dickson 1934], and [Bourbaki 1968, Chapter V, § 5, Ex. 6 on pp. 137–8] for a beautiful proof; the

[4]The best known examples of absolutely irreducible representations are the defining representations of the Weyl group of E_8 and the isometry group of the Leech lattice. Both of those representations are defined over $\mathbb{Q}$; thus the uniqueness up to scaling of the stable lattices for those groups is already contained in the work of Thompson [1976], who gave those examples as well as the 248-dimensional representation of his sporadic simple group. Gross's paper [Gross 1990] extends Thompson's work to several classes of representations not defined over $\mathbb{Q}$, and gives many examples.

Dickson invariants and the invariants of subgroups of $\mathrm{GL}_n(\mathbb{F}_q)$ are treated in greater detail in the last chapter of [Benson 1993].) In our case, $(q,n)=(2,3)$, so the degrees are $4,6,7$. Indeed (1.18) reduces mod 2 to[5] $\Phi_{21}^2=\Phi_{14}^3$, so that mod 2 there is a new invariant Φ_7 such that $\Phi_7^2=\Phi_{14}$, $\Phi_7^3=\Phi_{21}$; the Dickson invariants for $\mathrm{GL}_3(\mathbb{F}_2)$ are this Φ_7 together with Φ_4,Φ_6 — note that indeed the degrees $4,6,7$ are 2^3-2^2, 2^3-2^1, 2^3-2^0.

There is a unique prime $\wp_7=(\sqrt{-7}\,)$ of O_k above 7. The action of G on the 3-dimensional $\mathbb{F}_7$-vector space $L/\wp_7L$ is then the unique reduction of (V,ρ) in characteristic 7. Since $\beta\equiv\bar\beta \bmod \wp_7$ for all $\beta\in O_k$, the G-invariant Hermitian form on L reduces to a non-degenerate quadratic form on $L/\wp_7L$, which G must respect. Thus the image of our representation $G\to\mathrm{GL}_3(\mathbb{F}_7)$ is contained in the orthogonal group $\mathrm{SO}_3(\mathbb{F}_7)$ (not merely $\mathrm{O}_3(\mathbb{F}_7)$ because $\rho(G)\subset\mathrm{SL}(V)$ already in characteristic zero). But we already know a 3-dimensional representation of $G\cong\mathrm{PSL}_2(\mathbb{F}_7)$ in characteristic 7, namely the symmetric square $\mathrm{Sym}^2(V_2)$ of its defining representation. [Note that the matrix -1 in the center of $\mathrm{SL}_2(\mathbb{F}_7)$ acts on $\mathrm{Sym}^2(V_2)$ by multiplication by $(-1)^2=+1$, which is to say trivially, so we actually do obtain a 3-dimensional representation of the quotient group $\mathrm{PSL}_2(\mathbb{F}_7)$.] Moreover, this representation has an invariant quadratic form, namely the discriminant of a binary quadric, and G acts on $\mathrm{Sym}^2(V_2)$ by linear transformations of determinant 1. Thus we obtain a map $\mathrm{PSL}_2(\mathbb{F}_7)\to\mathrm{SO}_3(\mathbb{F}_7)$. The image is not quite all of $\mathrm{SO}_3(\mathbb{F}_7)$; indeed

$$\mathrm{SO}_3(\mathbb{F}_7)\cong\mathrm{Aut}(G). \tag{1.26}$$

Both groups have order $336=2\cdot168$, so to obtain the isomorphism (1.26) we need only extend the action of G on $\mathrm{Sym}^2(V_2)$ to $\mathrm{Aut}(G)\cong\mathrm{PGL}_2(\mathbb{F}_7)$. To do this, begin by choosing for each element of $\mathrm{Aut}(G)-G$ a representative $\gamma\in\mathrm{GL}_2(\mathbb{F}_7)$ of determinant -1; such a γ exists since -1 is not a square in $\mathbb{F}_7$, and is well-defined up to $\gamma\leftrightarrow-\gamma$. Then γ induces a linear transformation $\mathrm{Sym}^2\gamma$ $[\,=\mathrm{Sym}^2(-\gamma)\,]$ of determinant -1 on $\mathrm{Sym}^2(V_2)$ that preserves the quadratic form. We thus obtain a well-defined $-\mathrm{Sym}^2\gamma\in\mathrm{SO}_3(\mathbb{F}_7)$ not contained in the image of G. These elements, together with $\mathrm{Sym}^2\gamma$ for $\gamma\in G$, fill out all of $\mathrm{SO}_3(\mathbb{F}_7)$. (Geometrically, the actions of PGL_2 and SO_3 induce automorphisms of $\mathbb{P}^1$ and of a conic in $\mathbb{P}^2$ respectively, and the isomorphism (1.26) reflects the identification of the conic with $\mathbb{P}^1$ [Fulton and Harris 1991, p. 273].) We've seen that the G part of $\mathrm{SO}_3(\mathbb{F}_7)$ is obtained from the action of G on $L/\wp_7L$. But $\mathrm{Aut}(G)$ acts on L too, and since $\wp_7$ is Galois-invariant, the conjugate-linear automorphisms of L also act on $L/\wp_7L$.

We thus see that, as in the mod-2 case, each automorphism of $L/\wp_7L$ preserving the quadratic form lifts uniquely to an automorphism (possibly conjugate-

[5]It might be objected that we should not be using (1.18) because that equation relates the invariants of the Klein model. But that model still reduces well in characteristic 2; its only flaw there is that the field of definition is too large: $\mathbb{F}_8$ instead of $\mathbb{F}_2$. But this does not affect the structure of the $\overline{\mathbb{F}}_2$-ring of invariants.

linear and/or of determinant -1) of L. Moreover, L "explains" the sporadic isomorphism between $\mathrm{SL}_3(\mathbb{F}_2)$ and $\mathrm{PSL}_2(\mathbb{F}_7)$: these two linear groups are just the mod-2 and mod-7 manifestations of the isometries of L.[6]

The invariant quadratic form on $L/\wp_7 L$ can also be seen by reducing the ring of G-invariants mod $\wp_7$. As in the characteristic-2 case, there is a new invariant Φ_2, and here each of Φ_4, Φ_6, Φ_{14} is proportional to the appropriate power Φ_2^2, Φ_2^3, Φ_2^7 of this invariant quadric! Note that our formulas (1.11,1.22) for the quartic invariant in the S_4 and rational S_3 models both reduce mod $\wp_7$ to perfect squares, namely $(X^2+Y^2+Z^2)^2$ and $\big(X^2+Y^2+Z^2+3(XY+YZ+ZX)\big)^2$. Curiously, though, it is the S_4 form that is pertinent for $\Phi_4=\Phi_2^2$; that Φ_4 is also a square mod 7 in the rational S_3 model is not directly relevant. This is because the matrices (1.6,1.7) for $\rho(G)$ in the S_4 model are $\wp_7$-integral, while the matrices (1.21) in the rational S_3 model have denominators $\sqrt{-7}$ and even 7, and thus do not reduce well mod $\wp_7$. [For each odd prime power q, the full ring of invariants of the three-dimensional representations of $\mathrm{O}_3(\mathbb{F}_q)$, $\mathrm{PSL}_2(\mathbb{F}_q)$, and the three intermediate groups have been determined by Kemper [1996, Theorem 2.4(c)]. Of these five groups, only two have polynomial invariants, including $\mathrm{O}_3(\mathbb{F}_q)$ but not $\mathrm{PSL}_2(\mathbb{F}_q)$ of $\{\pm 1\}\times\mathrm{PSL}_2(\mathbb{F}_q)$. In our case of $q=7$, the invariants of $\mathrm{O}_3(\mathbb{F}_7)$ are generated by Φ_2, Φ_{21}^2, and a new invariant Φ_8 given by $X^8+Y^8+Z^8$ in the coordinates of the reduced S_4 model; G and $\pm G$ do not have polynomial invariant rings, though another index-2 subgroup of $\mathrm{O}_3(\mathbb{F}_7)$ has invariant ring $\mathbb{F}_7[\Phi_2,\Phi_8,\Phi_{21}]$. See [Kemper 1996] for further details.]

2. The Klein Quartic $\mathcal{X}$ as a Riemann Surface

2.1. The action of G on $\mathcal{X}$. The action of G on V induces an action on the projective plane $(V-\{\mathbf{0}\})/\mathbb{C}^*\cong\mathbb{CP}^2$, and on the Klein quartic $\mathcal{X}\subset\mathbb{CP}^2$, which is the zero-locus of the invariant quartic polynomial Φ_4. We use this to describe the geometry of $\mathcal{X}$.

We have seen already that Φ_4 is an irreducible polynomial. Thus its zero locus $\mathcal{X}$ is an irreducible curve. An irreducible plane quartic curve can have at most $\binom{4-1}{2}=3$ singularities. Any singular points of $\mathcal{X}$ would be permuted by G; since the largest proper subgroups of G have index 7, each singular point would have to be fixed by G. But G fixes no point on $\mathbb{CP}^2$ because the representation (V,ρ) is irreducible. Thus $\mathcal{X}$ has no singularities, so is a curve of genus 3 canonically embedded in $\mathbb{CP}^2$.

Since each element of G other than the identity can have only finitely many fixed points on $\mathcal{X}$, there are only a finite number of orbits of G of size less than $\#G=168$. We next describe these orbits and their point stabilizers:

[6]Several of the other sporadic isomorphisms between linear groups in different characteristics are likewise explained by highly symmetrical lattices in small dimension. For instance the Weyl group of E_6 occurs as both an orthogonal group acting on $\mathbb{F}_3^5$ and a symplectic group acting on $\mathbb{F}_2^6$, these vector spaces arising as $E_6/3E_6^*$ and $E_6/2E_6$. See [Kneser 1967].

PROPOSITION. (i) *Each of the eight 7-Sylow subgroups $H_7 \subset G$ has three fixed points in $\mathbb{CP}^2$ and is the stabilizer in G of each of these three points, all of which are on $\mathcal{X}$. The $8 \cdot 3 = 24$ points thus obtained are all distinct and constitute a single orbit of G. They are Weierstrass points of $\mathcal{X}$ of weight 1, and $\mathcal{X}$ has no other Weierstrass points.*

(ii) *Each of the twenty-eight 3-Sylow subgroups $H_3 \subset G$ has three fixed points in $\mathbb{CP}^2$. The normalizer $N(H_3)$ of H_3 in G, isomorphic with the symmetric group S_3, is the stabilizer in G of one of these points; this point is not on $\mathcal{X}$. The remaining fixed points of H_3 are on $\mathcal{X}$ and each has stabilizer H_3. The line joining these two points is the unique line of $\mathbb{CP}^2$ stable under $N(H_3)$, and is tangent to $\mathcal{X}$ at both points. The $28 \cdot 2 = 56$ points thus obtained are all distinct and constitute a single orbit of G. The lines joining pairs of these points with the same stabilizer are the 28 bitangents of $\mathcal{X}$.*

(iii) *Each of the twenty-one 2-element subgroups $H_2 \subset G$ fixes a point and a line in $\mathbb{CP}^2$. The normalizer $N(H_2)$ of H_2 in G, isomorphic with the 8-element dihedral group, is the stabilizer in G of the fixed point, which is not on $\mathcal{X}$. The fixed line meets $\mathcal{X}$ in four distinct points, each of which has stabilizer H_2 in G; these four points are permuted transitively by $N(H_2)$. The $21 \cdot 4 = 84$ points thus obtained are all distinct and constitute a single orbit of G.*

(iv) *Every G-orbit in $\mathcal{X}$, other than the orbits of size 24, 56, 84 described in (i), (ii), (iii) above, has size 168 and trivial stabilizer.*

PROOF. Since there are no points of $\mathbb{CP}^2$ fixed by all of G, the stabilizer of every point $P \in \mathbb{CP}^2$ must be contained in a maximal subgroup. For both kinds of maximal subgroup we have representations by monomial matrices relative to a suitable choice of coordinates, which let us readily describe the point stabilizers.

If the stabilizer $S(P)$ has even order it must be contained in one of the 24-element subgroups. In the coordinates of the S_4 model, we find that such a point P must be one of:

- a unit vector, with $S(P)$ an 8-element dihedral group;
- a vector $(1 : \pm 1 : \pm 1)$, with $S(P) \cong S_3$;
- a permutation of $(1 : \pm 1 : 0)$, with $S(P)$ a noncyclic group of order 4 (these last three cases coming from an opposite pair of faces, edges, or sides of the cube respectively);
- a permutation of $(1 : i : 0)$, with $S(P)$ a cyclic group of order 4, or
- a permutation of $(1 : x : \pm x)$ for some $x \notin \{0, \pm 1\}$, with $S(P)$ a two-element group.[7]

Moreover, the only nontrivial groups of odd order in S_4 are its 3-Sylows, which are conjugate to the group of cyclic permutations of the coordinates; this group

[7] There are several $x \notin \{0, \pm 1\}$ for which the stabilizer of this point in G is larger, but then that stabilizer is contained in a different maximal $S_4 \subset G$, and the point's coordinates in *that* subgroup's S_4 model appear earlier in this list.

fixes $(1:1:1)$, which we already saw has stabilizer S_3, and the two points $(1:e^{\pm 2\pi i/3}:e^{\mp 2\pi i/3})$. The stabilizer of each of these last points must be the 3-Sylow: it cannot be a larger subgroup of S_4, because we have already accounted for all of these; and the only other possibility would be a 21-element subgroup, which has no fixed points at all because it acts irreducibly on $\mathbb{CP}^2$. Turning to subgroups of the 21-element subgroup, we use the coordinates of the Klein model: the 7-element normal subgroup $\langle g\rangle$ fixes only the three unit vectors, and all 3-element subgroups are conjugate to $\langle h\rangle$ which fixes only $(1:1:1)$ and the two points

$$(1:e^{\pm 2\pi i/3}:e^{\mp 2\pi i/3}).$$

Clearly the first of these is also fixed by the involution (1.3). From our analysis of the S_4 model it follows that its stabilizer is the S_3 generated by h and that involution, while the other two fixed points of h have stabilizer $\langle h\rangle$.

Moreover, using the explicit formula for Φ_4 in the S_4 and Klein models we see that the stabilizers of any points of $\mathcal{X}$ must be cyclic of order 1,2,3, or 7. Thus part (iv) of the Proposition will follow from the first three parts.

Now a Weierstrass point of any Riemann surface of genus $w > 1$ is a point at which some holomorphic differential vanishes to order at least w. (See [Arbarello et al. 1985, 41–43] for the facts we'll need on Weierstrass points.) For a smooth plane quartic, the holomorphic differentials are linear combinations of the coordinates, so since $w = 3$ the Weierstrass points are those at which some line meets the curve at least triply, which is to say the inflection points of the curve. In our case the tangent to

$$\mathcal{X}: X^3Y + Y^3Z + Z^3X = 0$$

at $(1:0:0)$ is the line $Y = 0$, which indeed meets $\mathcal{X}$ triply at that point. Thus $(1:0:0)$ is a Weierstrass point, and by G-symmetry so are all 24 points in its orbit. But each Weierstrass point of a Riemann surface has a positive integral weight, and the sum of these weights is $w^3 - w$. Since this is 24 in our case, each point has weight 1 and there are no other Weierstrass points, as claimed. (Knowing the $w^3 - w$ formula we could have also concluded this directly from the existence of a unique orbit of size as small as 24, even without computing that it consists of inflection points.) We have thus proved Part (i) of the proposition.

(ii) First we check that $N(H_3)$ is indeed S_3. Since all 3-Sylows are conjugate in G, it is enough to do this when H_3 is contained in a maximal S_4. But the normalizer of every 3-element subgroup of S_4 is an S_3, so its normalizer in G is a subgroup of even order that is thus contained in a (perhaps different) maximal S_4, so is indeed S_3 as claimed.

To get at the fixed points of H_3 and its normalizer we again use the Klein model. We find that the fixed point $(1:1:1)$ of h is not on $\mathcal{X}$, while the other two fixed points are. Moreover the line connecting those two points is $X+Y+Z=0$; solving for Z and substituting into Φ_4 we obtain $-(X^2+XY+Y^2)^2$, so this line is indeed a bitangent of $\mathcal{X}$. That any smooth plane quartic curve has 28 bitangents is well known; see for instance [Hartshorne 1977, p. 305, Ex. 2.3h]. The remaining claims of (ii) either follow, as in (i), from the conjugacy in G of all 3-Sylow subgroups, or were already established during the above analysis of the stabilizers of points in $\mathbb{CP}^2$.

(iii) Again we first check that $N(H_2)$ is as claimed, using the fact that the involutions in G constitute a single conjugacy class. The normalizer of a double transposition in $S_4 \subset G$ is an 8-element dihedral group. Thus its normalizer in G is either that group, a maximal S_4, or all of G, but the last two are not possible because these groups have trivial centers. So $N(H_2)$ is indeed an 8-element dihedral group.

The noncyclic 4-group $N(H_2)/H_2$ acts on the fixed line of H_2 and on its intersection with $\mathcal{X}$. Since no point of $\mathcal{X}$ may have stabilizer properly containing H_2, the number of points of $\mathcal{X}$ on the fixed line must be a multiple of 4. But the intersection of a line with a smooth quartic curve consists of at least 1 and at most 4 points. Thus there are four fixed points of H_2 on $\mathcal{X}$, transitively permuted by $N(H_2)$. The remaining claims of (iii) follow as before. □

Corollary [Klein 1879b, § 6]. *The* 24-, 56- *and* 84-*point orbits are the zero loci of* Φ_6, Φ_{14}, *and* Φ_{21} *on* $\mathcal{X}$, *each with multiplicity* 1.

Proof. Since none of Φ_6, Φ_{14}, and Φ_{21} is a multiple of Φ_4, these polynomials do not vanish identically on $\mathcal{X}$, so their zero loci contain respectively 24, 56, and 84 points with multiplicity. Since the polynomials are G-invariant, their zero loci must be positive linear combinations of G-orbits. But by the Proposition there are only three orbits of size < 168. Moreover none of the integers 24, 56, 84 can be written as a nonnegative integer combination of the others: this is clear for 24, which is the smallest of the three; and almost as clear for 56, which is not a multiple of 24, and for 84, which is congruent to neither 0 nor 56 mod 24. Thus the vanishing loci can only be as claimed in the Corollary. □

(The Φ_6 case could also have been obtained from (1.13), since the inflection points of any smooth plane curve $P(X, Y, Z) = 0$ are the zeros of the Hessian $H(P)$ [Coolidge 1931, p. 95, Theorem 18]. The case of Φ_{21} could also be deduced from our description of Φ_{21} as the product of linear forms fixed by the 21 involutions in G. Klein also identifies the zeros of Φ_{21} on $\mathcal{X}$ with the curve's 84 "sextactic points", that is, the points at which the osculating conic meets $\mathcal{X}$ with multiplicity 6 rather than the generic 5.)

Hirzebruch [1983, pp. 120, 140] draws attention to the configuration in $\mathbb{CP}^2$ of the 21 lines fixed by involutions in G. Three of these meet at each of the 28 points fixed by subgroups $S_3 \subset G$, and four lines meet at each of the 21 points fixed by 2-Sylows in G. These are all the points of $\mathbb{CP}^2$ that lie on more than one of the 21 lines. In the notation of [Hirzebruch 1983], we thus have a configuration of $k = 21$ lines with $t_3 = 28$, $t_4 = 21$, and $t_n = 0$ for $n \neq 3, 4$. Thus this is one of the few nondegenerate line configuration known to achieve equality in the inequality

$$t_2 + \tfrac{3}{4}t_3 \geq k + \sum_{n>4}(n-4)t_n$$

of [Hirzebruch 1983, p. 140].

We can also use this Proposition to obtain, via the Riemann–Hurwitz formula, the genus of the quotient of $\mathcal{X}$ by each subgroup $H \subset G$: the quotient by the trivial group is of course $\mathcal{X}$ itself, with genus 3; the quotient by a cyclic subgroup of order 2, 3, or 4 is a curve of genus 1; and the quotient by any other subgroup has genus zero. Another way to obtain these is to identify the space $H_1(\mathcal{X}/H)$ of holomorphic differentials on $\mathcal{X}/H$ with the subspace $(H_1(\mathcal{X}))^H$ of such differentials on $\mathcal{X}$ fixed by H. But since $\mathcal{X}$ is a smooth plane quartic, we can identify $H_1(\mathcal{X})$ with the space of linear forms in the coordinates. Thus in our case the representation of G on $H_1(\mathcal{X})$ is isomorphic with (V^*, ρ^*), and we may recover the dimension of the subspace fixed by each subgroup H from the character table.

Since the quotient of $\mathcal{X}$ by the 7-Sylow $\langle g \rangle \subset G$ has genus 0, we can regard $\mathcal{X}$ as a cyclic cover of $\mathbb{CP}^1$ of degree 7. We can see this explicitly: the covering map sends $(X : Y : Z) \in \mathcal{X}$ to $(X^3Y : Y^3Z : Z^3X)$ on the line

$$\{(a : b : c) \in \mathbb{CP}^2 : a + b + c = 0\}. \tag{2.1}$$

Then $(Y/Z)^7 = ab^2/c^3$, and $(X : Y : Z)$ is determined by $(a : b : c)$ together with the seventh root Y/Z of ab^2/c^3. Thus if we take $y = -Y/Z$ and $x = b/c$ we find that $\mathcal{X}$ is birational with the curve

$$y^7 = x^2(x+1). \tag{2.2}$$

This model of $\mathcal{X}$ exhibits the action of the 21-element subgroup $\langle g, h \rangle$ of G: g multiplies y by ζ^{-1}, while g cyclically permutes a, b, c (or equivalently the points $-1, 0, \infty$ on the x-line). It also lets us write periods of differentials on $\mathcal{X}$ as linear combinations of Beta integrals. For instance, for the form dx/y^3 we find

$$\int_{-\infty}^{-1} \frac{dx}{y^3} = B\left(\tfrac{2}{7}, \tfrac{4}{7}\right), \quad \int_{-1}^{0} \frac{dx}{y^3} = B\left(\tfrac{1}{7}, \tfrac{4}{7}\right), \quad \int_{0}^{\infty} \frac{dx}{y^3} = B\left(\tfrac{1}{7}, \tfrac{2}{7}\right); \tag{2.3}$$

the identity $\Gamma(u)\Gamma(1-u) = \pi/\sin \pi u$ shows that each of these integrals is a K_+ multiple of

$$\Pi_7 := \frac{1}{\pi\sqrt{7}}\,\Gamma\left(\tfrac{1}{7}\right)\Gamma\left(\tfrac{2}{7}\right)\Gamma\left(\tfrac{4}{7}\right), \tag{2.4}$$

and thus that all the periods of dx/y^3 on $\mathcal{X}$ are in $K\Pi_7$. We later (2.12) use this to evaluate a complete elliptic integral as a multiple of Π_7.

We also compute for later use the quotient curve $\mathcal{X}/\langle h \rangle$ of genus 1. Since Φ_4 is not fixed by odd coordinate permutations, we can do this by multiplying Φ_4

by its image under such a permutation, and expressing the resulting symmetric function

$$(X^3Y + Y^3Z + Z^3X)(X^3Z + Z^3Y + Y^3X) \tag{2.5}$$

in terms of the elementary symmetric functions

$$s_1 = X + Y + Z, \quad s_2 = XY + YZ + ZX, \quad s_3 = XYZ. \tag{2.6}$$

We find that (2.5) is

$$s_2^4 + s_3(s_1^5 - 5s_1^3 s_2 + s_1 s_2^2 + 7s_1^2 s_3). \tag{2.7}$$

We thus get an affine model for $\mathcal{X}/\langle h\rangle$ by setting this polynomial equal to zero and substituting 1 for s_1:

$$7s_3^2 + (s_2^2 - 5s_2 + 1)s_3 + s_2^4 = 0. \tag{2.8}$$

To put this in Weierstrass form, divide (2.8) by s_2^4 and rewrite it as

$$7\left(\frac{s_3}{s_2^2}\right)^2 + (s_2^{-2} - 5s_2^{-1} + 1)\frac{s_3}{s_2^2} + 1 = 0. \tag{2.9}$$

Let $u = s_3/s_2^2$. Then (2.9) is a quadratic polynomial in s_2^{-1} over $\mathbb{Q}(u)$, so it has a root if and only if its discriminant $-28u^3 + 21u^2 - 4u$ is a square. The further substitution $u = -1/x$ then yields the desired form

$$E_k : y^2 = 4x^3 + 21x^2 + 28x \tag{2.10}$$

of the quotient curve. We can then compute that the curve has j-invariant $-3375 = -15^3$, and thus has complex multiplication (CM) by O_k. We note for future reference that the unit vectors, which have $s_2 = s_3 = 0$, map to the point at infinity of E_k, while the branch points of the cover $\mathcal{X}\to E_k$ are the fixed points $(1 : e^{\pm 2\pi i/3} : e^{\mp 2\pi i/3})$ of h, which have $s_1 = s_2 = 0$ and turn out to map to two points on E_k whose x-coordinates are roots $-\alpha$, $-\bar{\alpha}$ of

$$x^2 - x + 7 = 0. \tag{2.11}$$

The 2-element group $N(\langle h\rangle)/\langle h\rangle = \langle h, s\rangle/\langle h\rangle$ acts on E_k. Since $\mathcal{X}/\langle h, s\rangle$ has genus 0, the involution in $\langle h, s\rangle/\langle h\rangle$ must multiply the invariant differential on E_k by -1. Thus it is of the form $P \leftrightarrow P_0 - P$ for some $P_0 \in E_k$ (using the group law on E_k), and is determined by the image of a single point. We compute that s takes the unit vectors to points on $\mathcal{X}$ whose coordinates are proportional to the three roots of $u^3 - 7u^2 + 49$, and that these points map to the 2-torsion point $(0, 0)$ on E_k. Thus this point is P_0; in other words, the nontrivial element of $\langle h, s\rangle/\langle h\rangle$ acts on E_k by the involution that switches the point at infinity with $(0, 0)$ but is *not* translation by that 2-torsion point of E_k.

We further find that the curve E_k has conductor 49. (To see that the conductor is odd, note that the linear change of variable $y = 2y_1 + x$ puts E_k in the form $y_1^2 + xy_1 = x^3 + 5x^2 + 7x$ with good reduction at 2.) This conductor is small enough that we may locate the curve in the tables of elliptic curves dominated by modular curves compiled by Tingley et al. (the "Antwerp Tables" in [Birch and Kuyk 1975]) and Cremona [1992]: the curve is listed as 49A and 49-A1 respectively. We find there that E_k is literally a modular elliptic curve: it is not only dominated by, but in fact isomorphic with, $X_0(49)$. We shall later obtain this isomorphism from the identification of $\mathcal{X}$ with the modular curve $X(7)$. Likewise the fact that E_k has CM by O_k is no accident: we shall see that if there is a nonconstant map from $\mathcal{X}$ to an elliptic curve then the elliptic curve has CM by some order (subring of finite index) in O_k; equivalently, such a curve must be isogenous with E_k. (It is clear that conversely a curve isogenous with E_k admits such a map, since we have just constructed a nonconstant map from $\mathcal{X}$ to E_k itself.) For instance this must be true of the quotient of $\mathcal{X}$ by one of the 21 two-element subgroups of G. Since these subgroups are all conjugate in G, the resulting curves are isomorphic; in fact the reader may check (starting from the S_4 model of $\mathcal{X}$, in which several of these involutions are visible) that these elliptic curves are all $\bar{\mathbb{Q}}$-isomorphic with E_k.

An algebraic map from $\mathcal{X}$ to E_k can be used to pull back an invariant differential on E_k to $H_1(\mathcal{X})$. Thus the periods of E_k can be evaluated in terms of the Beta integrals that arise in the periods of $\mathcal{X}$. This yields the formula

$$\int_0^\infty \frac{dx}{\sqrt{4x^3+21x^2+28x}} = \tfrac{1}{4}\Pi_7 = \frac{1}{4\pi\sqrt{7}}\Gamma\left(\tfrac{1}{7}\right)\Gamma\left(\tfrac{2}{7}\right)\Gamma\left(\tfrac{4}{7}\right), \tag{2.12}$$

equivalent to Selberg and Chowla's result [1967, pp. 102–3]; its explanation via $\mathcal{X}$ is essentially the argument of Gross and Rohrlich [1978], though they pulled the differential all the way back to the Fermat curve $\mathcal{F}_7$, for which see Section 3.2 below.

2.2. $\mathcal{X}$ as the simplest Hurwitz curve. A classical theorem of Hurwitz ([1893]; see also [Arbarello et al. 1985, Chapter I, Ex. F-3 ff., pp. 45–47]) asserts that a Riemann surface S of genus $g>1$ can have at most $84(g-1)$ automorphisms, and a group of order $84(g-1)$ is the automorphism group of some Riemann surface of genus g if and only if it is generated by an element of orders 2 and one of order 3 such that their product has order 7. In that case the quotient of S by the group is the Riemann sphere, and the quotient map $S \to \mathbb{CP}^1$ is ramified above only three points of $\mathbb{CP}^1$, with the automorphisms of orders $2, 3, 7$ of S appearing as the deck transformations lifted from cycles around the three branch points. Thus the group elements of orders $2, 3, 7$ specify S by Riemann's existence theorem for Riemann surfaces. Note that the construction does not depend on the location of the three branch points on $\mathbb{CP}^1$, because $\mathrm{Aut}(\mathbb{CP}^1) = \mathrm{PGL}_2(\mathbb{C})$ acts on $\mathbb{CP}^1$ triply transitively.

A Riemann surface with the maximal number $84(g-1)$ of automorphisms, regarded as an algebraic curve over $\mathbb{C}$, is called a *Hurwitz curve* of genus g. Necessarily $g \geq 3$, because a curve C genus 2 over $\mathbb{C}$ has a hyperelliptic involution ι, and $\mathrm{Aut}(C)/\{1,\iota\}$ is the subgroup of $\mathrm{PGL}_2(\mathbb{C}) = \mathrm{Aut}(\mathbb{CP}^1)$ permuting the six ramified points, but the stabilizer in $\mathrm{Aut}(\mathbb{CP}^1)$ of a six-point set has size at most 24. So a Hurwitz curve must have genus at least 3. We know already that $\mathcal{X}$ is such a curve. In fact one may check that G is the only group of order 168 satisfying the Hurwitz condition, and that up to $\mathrm{Aut}(G)$ there is a unique choice of elements of orders $2, 3$ in G whose product has order 7. (For instance we may take the involution s and the 3-cycle sg.) Thus $\mathcal{X}$ is the unique Hurwitz curve of genus 3. We readily write the quotient map $\mathcal{X} \to \mathcal{X}/G \cong \mathbb{CP}^1$ explicitly, using our invariant polynomials $\Phi_6, \Phi_{14}, \Phi_{21}$: a point $(X : Y : Z)$ on $\mathcal{X}$ maps to

$$j := \frac{\Phi_{14}^3}{\Phi_6^7} = \frac{\Phi_{21}^2}{\Phi_6^7} + 1728 \tag{2.13}$$

on $\mathbb{CP}^1$. Note that this is a rational function of degree $4 \cdot 42 = 168 = \#G$ on $\mathcal{X}$, and thus of degree 1 on $\mathcal{X}/G$. That the two expressions in (2.13) are indeed equal on $\mathcal{X}$ follows from (1.19). We then see from (2.13) that the branch points of orders $2, 3, 7$ on $\mathbb{CP}^1$ have j coordinates $1728, 0, \infty$ respectively. Of course we have chosen this coordinate j on $\mathcal{X}/G \cong \mathbb{CP}^1$ to facilitate the identification of $\mathcal{X}$ and $\mathcal{X}/G$ with the modular curves $X(7)$ and $X(1)$ later in this paper.

Hurwitz curves can also be characterized in terms of their uniformization by the hyperbolic plane $\mathcal{H}$. Any Riemann surface S of genus > 1 can be identified with $\mathcal{H}/\pi_1(S)$; conversely, any discrete co-compact subgroup $\Gamma \subset \mathrm{Aut}(\mathcal{H}) \cong \mathrm{PSL}_2(\mathbb{R})$ that acts freely on $\mathcal{H}$ (that is, every point has trivial stabilizer) yields a Riemann surface $\mathcal{H}/\Gamma$ of genus > 1 whose fundamental group is Γ. The automorphism group of $\mathcal{H}/\Gamma$ is $N(\Gamma)/\Gamma$, where $N(\Gamma)$ is the normalizer of Γ in $\mathrm{Aut}(\mathcal{H})$. It follows that $\mathcal{H}/\Gamma$ is a Hurwitz curve if and only if $N(\Gamma)$ is the *triangle group* $G_{2,3,7}$ of orientation-preserving transformations generated by reflections in the sides of a given hyperbolic triangle with angles $\pi/2, \pi/3, \pi/7$ in $\mathcal{H}$. Equivalently, Γ is to be a normal subgroup of $G_{2,3,7}$. Since $G_{2,3,7}$ has the presentation

$$G_{2,3,7} = \left\langle \sigma_2, \sigma_3, \sigma_7 \mid \sigma_2^2 = \sigma_3^3 = \sigma_7^7 = \sigma_2\sigma_3\sigma_7 = 1 \right\rangle \tag{2.14}$$

(with σ_j being a $2\pi/j$ rotation about the π/j vertex of the triangle), this yields our previous characterization of the groups that can occur as $\mathrm{Aut}(S) = G_{2,3,7}/\Gamma$. In Section 4.4 we identify $\mathcal{X}$ with a Shimura modular curve by recognizing $G_{2,3,7}$ as an arithmetic group in $\mathrm{PSL}_2(\mathbb{R})$, and $\pi_1(\mathcal{X})$ with a congruence subgroup of $G_{2,3,7}$.

2.3. The Jacobian of $\mathcal{X}$. We have noted already that the representation of G on $H_1(\mathcal{X})$ isomorphic with (V^*, ρ^*). In particular, the representation is irreducible and defined over k, and its character takes values $\notin \mathbb{Q}$. It follows as in [Ekedahl and Serre 1993] that the Jacobian $J = J(\mathcal{X})$ is isogenous to the cube of an elliptic curve with CM by O_k. This does not determine J completely, but the fact that G acts on the period lattice of J means that this period lattice is proportional to L, and this does specify J. (See [Mazur 1986, pp. 235–6], where this is attributed to Serre; also compare [Buser and Sarnak 1994, Appendix 1], where the packing of congruent spheres in $\mathbb{R}^6$ obtained from L is conjectured to maximize the density of a packing coming from the period lattice of the Jacobian of a curve of genus 3.) In the notation of [Serre 1967] we have[8] $J \cong E_k \otimes L$.

We next describe a Mordell–Weil lattice associated with $\mathcal{X}$; see for instance [Elkies 1994] for more background on Mordell–Weil lattices.

Let E be an elliptic curve, and consider algebraic maps from $\mathcal{X}$ to E. These constitute an abelian group using the group law on E. This group may also be regarded as the group of rational points of E defined over the function field of $\mathcal{X}$; we thus call it the *Mordell–Weil group* M of maps from $\mathcal{X}$ to E, in analogy with the Mordell–Weil group of an elliptic curve over a number field. This group contains a subgroup isomorphic with E, namely the group of constant maps; the quotient group M/E may in turn be identified (via the embedding of X into J) with the group of morphisms from J to E. It follows that this group is trivial unless E has CM by an order in O_k, in which case it is a free abelian group of rank 6. This proves our earlier claim that the elliptic curves E admitting a nonconstant map from $\mathcal{X}$ are exactly the curves isogenous with E_k.

The function $\hat{h} : M \to \mathbb{Z}$ taking each $f : \mathcal{X} \to E$ to twice its degree as a rational map turns out to be a quadratic form. (For Riemann surfaces this is easy to see: let ω be a nonzero invariant differential on E; then $f \mapsto f^*\omega$ is a group homomorphism from M to $H_1(\mathcal{X})$, and $\hat{h}(f) = 2\deg(f)$ is the image of $f^*\omega$ under the quadratic form $\theta \mapsto 2\int_{\mathcal{X}} \theta \wedge \bar{\theta} / \int_C \omega \wedge \bar{\omega}$. Several proofs that $\hat{h}$ is a quadratic form valid in arbitrary characteristic are given in [Elkies 1994]. We use the notation $\hat{h}$ because this is a special case of the Néron–Tate canonical height; note that thanks to the factor of 2 the associated bilinear pairing

$$\langle f_1, f_2 \rangle = \tfrac{1}{2}\bigl(\hat{h}(f_1 + f_2) - \hat{h}(f_2) - \hat{h}(f_1)\bigr)$$

is integral.) This quadratic form is positive-definite on the free abelian group M/E, and gives this group the structure of a Euclidean lattice, which we thus call the *Mordell–Weil lattice* of maps from $\mathcal{X}$ to E.

Assume now that E is an elliptic curve with CM by O_k, i.e. that E is isomorphic with E_k. Then the Mordell–Weil lattice inherits the action of O_k on E as well as the action of G on $\mathcal{X}$. Therefore it is isomorphic with our lattice L^* of (1.25) up to scaling. Moreover, the quadratic form $\hat{h}$ satisfies the identity $\hat{h}(\beta f) = |\beta|^2 \hat{h}(f)$ for each $\beta \in O_k$, because $|\beta|^2$ is the degree of the isogeny $\beta : E \to E$. Thus $\hat{h}$ is a Hermitian pairing on L. This pairing is again unique up to scaling, this time because V^* is unitary and Hermitian (see again [Gross 1990]). If we identify L^* with the lattice generated by the three vectors (1.25) then we have

$$\hat{h}(v) = |v_1|^2 + |v_2|^2 + |v_3|^2. \tag{2.15}$$

[8]Serre actually defines $E \otimes L$ (or rather $L^* * E$) only when L is a lattice of rank 1 over $\mathrm{End}(E)$, but for each $g \geq 1$ the same construction for a lattice of rank g yields a polarized abelian variety isogenous with E^g.

This lattice has 21 pairs of vectors such as $(2,0,0)$ of minimal nonzero norm 4. These correspond to maps of degree 2 from C to E, which in turn are indexed by the 21 involutions $g \in G$. Each g is counted twice, because there are up to translation in E two ways to identify the quotient curve $\mathcal{X}/\{1,g\}$ with E, each yielding a map: $\mathcal{X} \to E$ of degree 2. Likewise the 28 pairs of vectors such as (α,α,α) of the next-lowest norm 6 correspond to maps of degree 3, all of which turn out to be quotient maps by the twenty-eight 3-Sylow subgroups of G. For each n the number N_n of maps of degree n up to translation on E is the number of vectors of norm $2n$ in L, which is the q^n coefficient of the theta series

$$\theta_L := \sum_{n=0}^{\infty} N_n q^n = \sum_{v \in L} q^{\frac{1}{2}\hat{h}(v)} \tag{2.16}$$

of L. But θ_L is a modular form of weight 3 with quadratic character on $\Gamma_0(7)$ fixed by the Fricke involution w_7 ([Gross 1990, § 9]; we shall encounter $\Gamma_0(7)$ and w_7 again in Section 4.2), and the space of such modular forms is 2-dimensional. The constraints $N_0 = 1$, $N_1 = 0$ determine θ_L uniquely, and we find

$$\begin{aligned}\theta_L &= \Bigl(\sum_{\beta \in O_k} q^{\beta\bar{\beta}}\Bigr)^3 - 6q \prod_{n=1}^{\infty} (1-q^n)^3 (1-q^{7n})^3 \qquad (2.17)\\ &= 1 + 42q^2 + 56q^3 + 84q^4 + 168q^5 + 280q^6 + 336q^7 + 462q^8 + \cdots .\end{aligned}$$

This confirms our values $N_2 = 42$ and $N_3 = 56$ and lets us easily calculate as many N_n as we might reasonably desire.

3. Arithmetic Geometry of $\mathcal{X}$

3.1. Rational points on $\mathcal{X}$. Faltings' theorem (né Mordell's conjecture) asserts that a curve of genus at least 2 over a number field has finitely many rational points. Unfortunately both of Faltings' proofs of this [1983; 1991] are ineffective, in that neither yields an algorithm for provably listing all the points; even for a specific curve of low genus over $\mathbb{Q}$ this problem can be very difficult. (See for instance [Poonen 1996].) Fortunately the special case of $\mathcal{X}$ is much easier. One shows that the elliptic curve E_k has rank zero, and its only rational points are the point at infinity and $(0,0)$. Since $\mathcal{X}$ admits a nonconstant map to E_k defined over $\mathbb{Q}$, namely the quotient map $\mathcal{X} \to \mathcal{X}/\langle h \rangle \cong E_k$, the $\mathbb{Q}$-rational points of $\mathcal{X}$ are just the rational preimages of the two points of $E_k(\mathbb{Q})$. We find that the only points of $\mathcal{X}(\mathbb{Q})$ are the obvious ones at $(1:0:0)$, $(0:1:0)$, $(0:0:1)$. Equivalently, the only integer solutions of $X^3Y + Y^3Z + Z^3X = 0$ are those in which at least two of the three variables vanish. This is all for the Klein model; one may likewise analyze the rational S_3 model for $\mathcal{X}$, computing[9] that its quotient by $\langle h \rangle$ is isomorphic with E_k, and that neither of the rational points

[9] This computation begins in the same way as our derivation of (2.10), but yields an equation $y^2 = x^4 - 10x^3 + 27x^2 - 10x - 27$ for the quotient curve; to bring this to Weierstrass form,

of E_k lies under a rational point of $\mathcal{X}$. However, the fact that $\mathcal{X}$ has no rational points in the rational S_3 model can be obtained much more simply, without any computation of quotient curves and analysis of elliptic curves over $\mathbb{Q}$: one need only observe that the polynomial (1.22) does not vanish mod 2 unless X, Y, Z are all even.

The proof that $E_k(\mathbb{Q})$ consists only of the point at infinity and $(0, 0)$ is an application of Fermat's method of descent. Suppose that $x \neq 0$ is a rational number such that $x(4x^2 + 21x + 28) = y^2$ for some $y \in \mathbb{Q}$. Necessarily $x > 0$, because $4x^2 + 21x + 28 > 0$ for all $x \in \mathbb{R}$. Write x as a fraction m/n in lowest terms. If x works then so does $7/x$ (note that $(7/x, -7y/x^2)$ is the translate of (x, y) by the 2-torsion point $(0, 0)$ in the group law of E_k). Replacing x by $7/x$ if necessary, we may assume that the exponents of 7 in the factorizations of m, n are both even. Then the integer $(n^2y)^2 = mn(4m^2 + 21mn + 28n^2)$ is a perfect square, and its factors m, n, $4m^2 + 21mn + 28n^2$ are relatively prime in pairs except possibly for common factors of $2 \cdot 49^r$ or $4 \cdot 49^r$. Thus either all three are squares, or one is a square and the each of the other two is twice a square. We claim that the latter is impossible. Indeed, since m, n cannot both be even, we would have either $(m, n) = (M^2, 2N^2)$ or $(m, n) = (2M^2, N^2)$. In the first case,

$$4m^2 + 21mn + 28n^2 = 2(2M^4 + 21M^2N^2 + 56N^4). \tag{3.1}$$

But M is odd (else m, n are both even), so $2M^4 + 21M^2N^2 + 56N^4$ is either 2 or 3 mod 4 according as N is even or odd; in neither case can it be a perfect square. In the second case, N is odd and

$$4m^2 + 21mn + 28n^2 = 2(8M^4 + 21M^2N^2 + 14N^4). \tag{3.2}$$

Again the parenthesized factor is either 2 or 3 mod 4, this time depending on the parity of M, so it cannot be a square.

So we conclude that m, n are both squares. Thus $x = x_1^2$ for some $x_1 \in \mathbb{Q}^*$, and $4x_1^4 + 21x_1^2 + 28 \in \mathbb{Q}^{*2}$. We "complete the square" by writing $\sqrt{4x_1^4 + 21x_1^2 + 28}$ as $2x^2 + (21 - \xi)/4$, finding

$$16\xi x^2 = \xi^2 - 42\xi - 7. \tag{3.3}$$

Necessarily $\xi \neq 0$ because the right-hand side has irrational roots. Thus we obtain a point on the elliptic curve

$$E_k' : \eta^2 = \xi(\xi^2 - 42\xi - 7) \tag{3.4}$$

other than the origin and $(0, 0)$. We then mimic the argument in the previous paragraph to show that either ξ or $-7/\xi$ must be a square. This time the possibility that must be excluded is that that one of them is $-\xi_1^2$ for some $\xi_1 \in \mathbb{Q}$. Taking $\xi_1 = M/N$, we would then have a square of the form $7N^4 - 42M^2N^2 - M^4$. But this is congruent to 3 mod 4 if either M or N is even, and to -4 mod 16

complete the square as we do several times in the sequel, for instance when obtaining (3.4) or in the calculation starting with (3.9).

if they are both odd, so again we reach a contradiction. Thus $\xi = \xi_1^2$ and we find that $\xi_1^4 - 42\xi_1^2 - 7$ is a square, say $\bigl(\xi^2 - (8x_2 + 21)\bigr)^2$. This yields $x_2\xi^2 = 4x_2^2 + 21x_2 + 28$; again the right-hand side has irrational roots, so we find $x_2 \in \mathbb{Q}^*$ such that $x_2(4x_2^2 + 21x_2 + 28) \in \mathbb{Q}^2$ — which is to say, a new point on E_k! Moreover, we can compute our original x or $7/x$ as a rational function in x_2 of degree 4, which means that if the numerator and denominator of x are at all large ($|M|, |N| > 100$ is more than enough) then those of x_2 are smaller. Iterating this descent process enough times, we eventually find a rational solution of $y^2 = 4x^3 + 21x^2 + 28x$ with nonzero $x = M/N$ such that $|M|, |N| \le 100$. But a direct search shows that there is no such x. This completes the proof that the only rational points on E_k are the two torsion points already known.

[In modern terminology, Fermat's method is "descent via a 2-isogeny" $E_k' \to E_k$ [Silverman 1986, pp. 301 ff.]. The method can be used on any elliptic curve with a rational 2-torsion point, and will often prove that the curve has only finitely many rational points. The reappearance of E_k at the second step, which makes it possible to iterate the process until reaching a small point, is due to the existence of a dual isogeny $E_k \to E_k'$ also of degree 2. Composing these two isogenies yields the multiplication-by-2 map on E_k; thus we proved in effect that any rational point on E_k is in either divisible by 2 or of the form $2P+(0,0)$ in $E_k(\mathbb{Q})$, and then used the fact that multiplication by 2 in $E_k(\mathbb{Q})$ quadruples the height to reduce the determination of $E_k(\mathbb{Q})$ to a finite search. In our case the 2-isogenous curve E_k' has j-invariant 255^3 and CM by $\mathbb{Z}[\sqrt{-7}\,]$; it is the elliptic curve numbered 49B in [Birch and Kuyk 1975] and 49-A2 in [Cremona 1992].]

It remains to find the preimages on $\mathcal{X}$ of the two rational points of E_k. We saw already that the point at infinity comes from the unit vectors on $\mathcal{X}$, and that the 2-torsion point $(x, y) = (0, 0)$ is the image of an $\langle h \rangle$-orbit of points on $\mathcal{X}$ whose coordinates are proportional to the three roots of $u^3 - 7u^2 + 49$. These roots (and their ratios) are contained in K_+ but not in $\mathbb{Q}$. Thus the unit vectors are the only rational points on $\mathcal{X}$, as claimed.

3.2. Fermat's Last Theorem for exponent 7. The Fermat curve

$$\mathcal{F}_7 : A^7 + B^7 + C^7 = 0$$

admits a nonconstant map to $\mathcal{X}$ defined over $\mathbb{Q}$, namely

$$(A : B : C) \mapsto (A^3C : B^3A : C^3B).$$

(The map is a cyclic unramified cover of degree 7, but we do not need this for now.) Thus any rational point on $\mathcal{F}_7$ maps to a rational point on $\mathcal{X}$. Having just listed the rational points on $\mathcal{X}$ we can thus determine the rational points on $\mathcal{F}$. It turns out that each point of $\mathcal{X}(\mathbb{Q})$ lies under a unique point of $\mathcal{F}(\mathbb{Q})$. This yields a proof of the case $n = 7$ of "Fermat's Last Theorem", a proof that is elementary in that it uses only tools available to Fermat (algebraic manipulation and 2-descent on an elliptic curve with a rational 2-torsion point); in particular it does not require arithmetic in cyclotomic number fields such as K. Indeed the proof is analogous to Fermat's own proof of the case $n = 4$, in the sense that in both cases one maps $\mathcal{F}_n$ to an elliptic curve and proves that the elliptic curve has rank 0; it is arguably easier than Euler's proof of the case $n = 3$, for which $\mathcal{F}_3$ is already an elliptic curve but the determination of $\mathcal{F}_3(\mathbb{Q})$ requires what we now recognize as a 3-descent. As is the case for $n = 4$, the map from $\mathcal{F}_7$ to E_k is a quotient map, here by a 21-element subgroup of $\mathrm{Aut}(\mathcal{F}_7)$ isomorphic with $\langle g, h\rangle \subset G$.

Stripped of all algebro-geometric machinery, this elementary proof runs as follows: Suppose there existed nonzero integers a, b, c such that $a^7 + b^7 + c^7 = 0$. Then

$$x := a^3c, \quad y := b^3a, \quad z := c^3b \tag{3.5}$$

would be nonzero integers with

$$x^3y + y^3z + z^3x = a^3b^3c^3(a^7 + b^7 + c^7) = 0, \tag{3.6}$$

which we showed impossible in the previous section.

Curiously there is yet another proof of the $n = 7$ case of Fermat along the same lines, which was discovered in the mid-19th century [Genocchi 1864][10] but is practically unknown today. Here we use the quotient of $\mathcal{F}_7$ by the group S_3 of coordinate permutations. This yields the following nice generalization of Fermat for $n = 7$:

THEOREM [Genocchi 1864][11]. *Let a, b, c be the solutions of a cubic $x^3 - px^2 + qx - r = 0$ with rational coefficients p, q, r. If $a^7 + b^7 + c^7 = 0$ then either $abc = 0$ or $a^3 = b^3 = c^3$.*

That is, the only rational points on $\mathcal{F}_7/S_3$ are the orbits of $(1 : -1 : 0)$ and $(1 : e^{2\pi i/3} : e^{-2\pi i/3})$. We compute equations for $\mathcal{F}_7/S_3$ by writing $a^7 + b^7 + c^7$ as a polynomial in the elementary symmetric functions

$$p = a + b + c, \qquad q = ab + ac + bc, \qquad r = abc \tag{3.7}$$

of a, b, c.

PROOF. We easily calculate

$$0 = a^7 + b^7 + c^7 = p^7 - 7p^5q + 7p^4r + 14p^3q^2 - 21p^2qr - 7pq^3 + 7pr^2 + 7q^2r \tag{3.8}$$

(for instance by using the fact that the power moments $\pi_n = a^n + b^n + c^n$ satisfy the recursion $\pi_{n+3} - p\pi_{n+2} + q\pi_{n+1} - r\pi_n = 0$ and starting from $\pi_0 = 3$, $\pi_1 = p$,

[10] From Dickson [1934, p. 746], footnote 85. Dickson further notes that Genocchi's method may be viewed as a simplification of Lamé's, and that Genocchi does not carry out the descent for proving that (3.9) has no finite rational points. We likewise leave the 2-descent on the equivalent curve (3.12) to the reader, who may either do it by hand using the method described in [Silverman 1986, pp. 301 ff.] or automatically with Cremona's mwrank program.

[11] In fact Genocchi states at the end of his paper that he had announced the results several years earlier in "*Cimento* di Torino, vol. VI, fasc. VIII, 1855."

$\pi_2 = p^2 - 2q$ to reach the formula (3.8) for π_7). Now if $p = 0$ then (3.8) reduces to $\pi_7 = 7q^2r$, so if $\pi_7 = 0$ then either $r = 0$ or $q = 0$, which yields $abc = 0$ or $a^3 = b^3 = c^3$ respectively. If on the other hand $p \neq 0$ then we may assume $p = 1$ by replacing a, b, c by $a/p, b/p, c/p$. We then find that (3.8) is a quadratic polynomial in r of discriminant $49q^4 - 98q^3 + 147q^2 - 98q + 21$. We note that the resulting elliptic curve

$$u^2 = 49q^4 - 98q^3 + 147q^2 - 98q + 21 \tag{3.9}$$

has rational points at infinity, and use them to obtain a Weierstrass form for the curve by the usual device of completing the square: let

$$u = 7(q^2 - q + 1 - 2t) \tag{3.10}$$

in (3.9) to find

$$7t(q^2 - q) = 7t^2 - 7t + 1, \tag{3.11}$$

a quadratic in q with discriminant $196t^3 - 147t^2 + 28t$. Thus $196t^3 - 147t^2 + 28t$ must be a square. Taking $t = -x/7$, then, we obtain the elliptic curve

$$-7y^2 = 4x^3 + 21x^2 + 28x, \tag{3.12}$$

which we recognize as the $\sqrt{-7}$-twist of E_k. Since that curve has CM by $\mathbb{Z}[\alpha]$, this new curve (3.12) is also 7-isogenous with E_k, and thus has rank zero. (This curve appears as 49C in [Birch and Kuyk 1975] and 49-A3 in [Cremona 1992].) In fact we can apply a 2-descent directly to (3.12) using the 2-torsion point $(0, 0)$, and then find that this point is the only rational point of (3.12) other than the point at infinity. But if $x = 0$ then $t = 7x = 0$ and (3.11) becomes $0 = 1$, which is impossible (indeed the points $x = 0, \infty$ on (3.12) come from the solutions $p = r = 0$ and $p = q = 0$ of (3.8) — which solution is which depends on the choice of square root u implicit in (3.9)). Thus indeed $p = 0$ in any rational solution of (3.8), which completes the proof of Genocchi's theorem. $\square$

[Along these lines we note that Gross and Rohrlich [1978] have shown that the orbits of $(1 : -1 : 0)$ and $(1 : e^{2\pi i/3} : e^{-2\pi i/3})$ also contain the only points of $\mathcal{F}_7$ rational over any number field of degree at most 3.]

3.3. Reduction of $\mathcal{X}$ modulo 2,3,7. For each of the primes $p = 2, 3, 7$ dividing $\#G$, the reduction of $\mathcal{X} \bmod p$ enjoys some remarkable extremal properties: maximal or minimal numbers of points over finite fields in each case, and maximal group of automorphisms for $p = 3$. We consider these three primes in turn.

Characteristic 2. Since we want all the automorphisms of G to be defined over $\mathbb{F}_2$, we use the S_4 or rational S_3 model for $\mathcal{X}$. Then the Jacobian of $\mathcal{X}$ is $\mathbb{F}_2$-isogenous to the cube of an elliptic curve with CM by $\mathbb{Z}[\alpha]$ and trace 1. It follows that the characteristic polynomial of Frobenius for $\mathcal{X}/\mathbb{F}_2$ is $(T^2-T+2)^3$, with triple roots $-\alpha, -\bar{\alpha}$. Thus for each $m \geq 1$ our curve has

$$2^m + 1 - 3\big((-\alpha)^m + (-\bar{\alpha})^m\big) \tag{3.13}$$

rational points over $\mathbb{F}_{2^m}$. We tabulate this for the first few m:

m	1	2	3	4	5	6	7	8	...
$\#(\mathcal{X}(\mathbb{F}_{2^m}))$	0	14	24	14	0	38	168	350	...

(3.14)

We noted already that the reduction mod 2 of the rational S_3 model for $\mathcal{X}$ has no $\mathbb{F}_2$-rational points. That it has no $\mathbb{F}_{32}$-rational points is rather more remarkable. By the Weil estimates, a curve of genus w over $\mathbb{F}_q$ has at least $q - 2wq^{1/2} + 1$ rational points; if $w > 1$, this lower bound may be negative, but only for $q \leq 4w^2 - 3$. In our case $w = 3$, this bound on q is 33, which is not a prime power, so $\mathbb{F}_{32}$ is the largest finite field over which a curve of genus 3 may fail to have any rational point. [For $w = 2$, the bound $4w^2 - 3$ is the prime 13, but Stark showed ([1973]; see in particular pages 287–288) that there is no pointless curve of genus 2 over $\mathbb{F}_{13}$; an explicit such curve over $\mathbb{F}_{11}$ is $y^2 = -(x^2+1)(x^4+5x^2+1)$.]

The 14 points of our curve over $\mathbb{F}_4$ are all the points of $\mathbb{P}^2(\mathbb{F}_4) - \mathbb{P}^2(\mathbb{F}_2)$. It is known that this is the maximal number of points of a genus-3 curve over $\mathbb{F}_4$ [Serre 1983a; 1983b; 1984]. Note that the only $\mathbb{F}_{16}$-points are those already defined over $\mathbb{F}_4$; indeed one can use the "Riemann hypothesis" (which is a theorem of Weil for curves over finite fields) to show as in [Serre 1983b] that a genus-3 curve over $\mathbb{F}_4$ with more than 14 points would have *fewer* $\mathbb{F}_{16}$ points than $\mathbb{F}_4$ points, and thus prove that 14 is the maximum. The 24 points over $\mathbb{F}_8$ likewise attain the maximum for a genus-3 curve over that field [Serre 1983a; 1984]. Note that the only $\mathbb{F}_{64}$-points are those already rational over a subfield $\mathbb{F}_4$ or $\mathbb{F}_8$.

Upon reading a draft of this paper, Serre noted that in fact for $m = 2, 3, 5$ the curve $\mathcal{X}$ is the unique curve of genus 3 over $\mathbb{F}_{2^m}$ with the maximal ($m = 2, 3$) or minimal ($m = 5$) number of rational points. He shows this as follows. Let $C/\mathbb{F}_{2^m}$ be any curve with the same number of points as $\mathcal{X}$. First Serre proves that C has the same eigenvalues of Frobenius as $\mathcal{X}$. For $m = 3, 5$ this follows from the fact that C attains equality in the refined Weil bound

$$\big|\#C(\mathbb{F}_q) - (q+1)\big| \leq g\lfloor 2q^{1/2}\rfloor \tag{3.15}$$

(see [Serre 1983b, Theorem 1]). For $m = 2$ we instead use the fact that $\#C(\mathbb{F}_{16}) \geq \#C(\mathbb{F}_4) = 14$. Serre then notes that in each of the three cases $m = 2, 3, 5$ we have $\alpha^m = (x \pm \sqrt{-7})/2$ for some $x \in \mathbb{Z}$, from which it follows that $\mathbb{Z}[(-\alpha)^m]$ is the full ring of integers in k. Thus the Jacobian of C is isomorphic as a principally polarized abelian threefold $E_k \otimes M$, where M is some indecomposable positive-definite unimodular Hermitian O_k-lattice of rank 3. But by Hoffmann's result [1991, Theorem 6.1] cited above, L is the unique such lattice. Thus C has the same Jacobian as $\mathcal{X}$, from which $C \cong \mathcal{X}$ follows by Torelli.

Since k has unique factorization, the condition $\alpha^m = \left(x \pm \sqrt{-7}\,\right)/2$ is equivalent to the Diophantine equation

$$x^2 + 7 = 2^n \tag{3.16}$$

with $n = m + 2$. (This equation also arises in [Serre 1983a], in connection with curves of genus 2 over $\mathbb{F}_{2^m}$ with many points, and even in coding theory [MacWilliams and Sloane 1977, p. 184], because it is equivalent to the condition that the volume of the Hamming sphere of radius 2 in $(\mathbb{Z}/2)^{(x-1)/2}$ be a power of 2.) Ramanujan observed[12] that, in addition to the cases $n = 3, 4, 5, 7$ already encountered, this equation has a pair of solutions $(n, x) = (15, \pm 181)$. We find that $(-\alpha)^{13}$ has negative real part, and conclude from Serre's argument that $\mathcal{X}$ is the unique curve of genus 3 over $\mathbb{F}_{2^{13}}$ with the maximal number of rational points, namely $8736 = 2^5 3 \cdot 7 \cdot 13$. Nagell [1960] was apparently the first to show that the Diophantine equation (3.16) has no further integer solutions.

The 24 points over $\mathbb{F}_8$ are, as could be expected, the reduction mod 2 of the 24-point orbit of Weierstrass points of $\mathcal{X}$ in characteristic zero. The $\mathbb{F}_4$ points require some more comment: since G acts on $\mathcal{X}$ by automorphisms defined over the prime field, it permutes these 14 points, whereas in characteristic zero there was no orbit as small as 14 in the action of G on $\mathcal{X}$, or even on $\mathbb{P}^2$. But in characteristic 2 the 24-element subgroups of $G \cong \mathrm{SL}_3(\mathbb{F}_2)$ arise naturally as stabilizers of points and lines in $\mathbb{P}^2(\mathbb{F}_2)$. The stabilizer of a line $\mathbb{P}^1(\mathbb{F}_2) \subset \mathbb{P}^2(\mathbb{F}_2)$ permutes the two points of the line rational over $\mathbb{F}_4$ but not $\mathbb{F}_2$; the subgroup fixing each of those points thus has index 2 in the line stabilizer. Moreover each point of $\mathbb{P}^2(\mathbb{F}_4) - \mathbb{P}^2(\mathbb{F}_2)$ lies on a unique $\mathbb{F}_2$-rational line. Thus the stabilizer of each of these points is a subgroup $A_4 \subset G$. Such a subgroup contains three involutions, each now having two instead of four fixed points on $\mathcal{X}$, and four 3-Sylows. Thus the 14 points of $\mathcal{X}(\mathbb{F}_4)$ are the reductions mod 2 of both the 56-point and the 84-point G-orbits. All points of $\mathcal{X}$ not defined over $\mathbb{F}_4$ or $\mathbb{F}_8$ have trivial stabilizer in G; such points first occur over $\mathbb{F}_{2^7}$, where the 168 points of $\mathcal{X}(\mathbb{F}_{2^7})$ constitute a single G-orbit. The image of this orbit, together with those of $\mathcal{X}(\mathbb{F}_4)$ and $\mathcal{X}(\mathbb{F}_8)$, account for the three $\mathbb{F}_2$-points of $\mathcal{X}/G \cong \mathbb{P}^1$. The $350 - 14 = 336$ points in $\mathcal{X}(\mathbb{F}_{2^8}) - \mathcal{X}(\mathbb{F}_{2^2})$ are likewise the preimages of the two points of $\mathcal{X}/G$ defined over $\mathbb{F}_4$ but not $\mathbb{F}_2$.

We conclude the description of $\mathcal{X}$ in characteristic 2 with an amusing observation of Seidel concerning the $\mathbb{F}_8$-rational points of $\mathcal{X}$, reported by R. Pellikaan at a 1997 conference talk. Since $\mathbb{F}_8$ is the residue field of the primes above 2 of K, the reductions mod 2 of the Klein and S_4 models of $\mathcal{X}$ become isomorphic over

[12] On page 120 of the *Journal of the Indian Mathematical Society*, Volume 5 #3 (6/1913), we find under "Questions for Solution":

> 464. (S. Ramanujan):—$(2^n - 7)$ is a perfect square for the values 3, 4, 5, 7, 15 of n. Find other values.

No other values were found, but it does not seem that a proof that none exist ever appeared in the *Journal*.

that field; we use the Klein model. Consider the $24 - 3 = 21$ points of $\mathcal{X}(\mathbb{F}_8)$ other than the three unit vectors. These may be identified with the solutions in $\mathbb{F}_8^*$ of the affine equation $x^3y + y^3 + x = 0$ (with $x = X/Z$, $y = Y/Z$) for the Klein model of $\mathcal{X}$. We choose (α) for our prime above 2, so ζ reduces to a root of $\zeta^3 + \zeta + 1$ in $\mathbb{F}_8$. The 21 solutions (x, y) are then entered in the following table:

x / y	1	ζ^3	ζ^6	ζ^2	ζ^5	ζ^1	ζ^4
1				•		•	•
ζ	•				•		•
ζ^2	•	•				•	
ζ^3		•	•				•
ζ^4	•		•	•			
ζ^5		•		•	•		
ζ^6			•		•	•	

(3.17)

(note that we have listed the x- and y-coordinates in different orders so as to make the $\langle g \rangle$ symmetry visible). Seidel's observation is that (3.17) is the adjacency matrix for the finite projective plane of order 2! The explanation is that for $x, y \in \mathbb{F}_8^*$,

$$x^3y + y^3 + x = 0 \iff x^4y^2 + xy^4 + x^2y = 0 \iff \operatorname{Tr}_{\mathbb{F}_8/\mathbb{F}_2} x^2y = 0. \quad (3.18)$$

Now $(x, y) \mapsto \operatorname{Tr}_{\mathbb{F}_8/\mathbb{F}_2} x^2y$ is a nondegenerate pairing from $\mathbb{F}_8 \times \mathbb{F}_8$ to $\mathbb{F}_2$, so if we regard $x \in \mathbb{F}_8$ an element of 3-dimensional vector space over $\mathbb{F}_2$ then y is a functional on that vector space and (3.18) is the condition that a nonzero functional annihilate a nonzero vector. Thus if we regard $x, y \in \mathbb{F}_8^*$ as 1- and 2-dimensional subspaces of $\mathbb{F}_2^3$ then $x^3y + y^3 + x = 0$ if and only if the x-line is contained in the y-plane, which is precisely the incidence relation on the points and lines of the finite projective plane $\mathbb{P}^2(\mathbb{F}_2)$.

Characteristic 3. Since 3 is inert in k, the smallest field over which all the automorphisms in G might be defined is $\mathbb{F}_9$. Again we make sure that they are in fact defined over that field by using the S_4 or rational S_3 model for $\mathcal{X}$. That 3 does not split in k also makes the elliptic curve E_k, with CM by O_k, supersingular in characteristic 3; we find that its characteristic polynomial of Frobenius over $\mathbb{F}_9$ is $(T+3)^2$, and hence that $\mathcal{X}$ has $9^m - 6(-3)^m + 1$ rational points over $\mathbb{F}_{9^m}$. Thus, depending on whether m is odd or even, $\mathcal{X}$ has the maximal or minimal number of rational points for a curve of genus 3 over $\mathbb{F}_{9^m}$. Moreover, the curve has 28 points over both $\mathbb{F}_9$ and $\mathbb{F}_{81}$, and thus maximizes the genus w of a curve $C/\mathbb{F}_9$ that can attain the Weil upper bound $9 + 6w + 1$ on $\#C(\mathbb{F}_9)$.

In fact this turns out to be a special case of a known construction of curves attaining the Weil bound over $\mathbb{F}_{q^2}$. Note that Φ_4, as given by either (1.11) or (1.22), reduces mod 3 to $X'^4+Y'^4+Z'^4$ or $A^4+B^4+C^4$. That is, *the Klein and Fermat quartics are isomorphic in characteristic 3.* Now for each prime power q, the equation $x^{q+1}+y^{q+1}+z^{q+1}=0$ defining the Fermat curve $\mathcal{F}_{q+1}$ can be written as

$$x^q x + y^q y + z^q z = 0. \tag{3.19}$$

For $a \in \mathbb{F}_{q^2}$ we note that $a^q a$ is just the norm of a from $\mathbb{F}_{q^2}$ to $\mathbb{F}_q$. This lets us easily count the solutions of (3.19) in $\mathbb{F}_{q^2}$, and we calculate that $\mathcal{F}_{q+1}$ has q^3+1 rational points over $\mathbb{F}_{q^2}$. Since this curve has genus $(q^2-q)/2$, it thus attains the Weil bound. Therefore its characteristic polynomial of Frobenius over $\mathbb{F}_{q^2}$ is $(T+q)^{q^2-q}$, so the number of $\mathbb{F}_{q^4}$-rational points of $\mathcal{F}_{q+1}$ is

$$q^4-(q^2-q)q^2+1=q^3+1 \tag{3.20}$$

again. If there were a curve $C/\mathbb{F}_{q^2}$ of genus $w>(q^2-q)/2$ attaining the Weil bound, its number $q^2+2qw+1$ of $\mathbb{F}_{q^2}$-rational points would exceed the number q^4-2q^2w+1 of points rational over $\mathbb{F}_{q^4}$; thus again $\mathcal{F}_{q+1}$ is the curve of largest genus attaining the Weil bound over $\mathbb{F}_{q^2}$. These properties of $\mathcal{F}_{q+1}$ over $\mathbb{F}_{q^2}$ are well-known, see for instance [Serre 1983a; 1984].

Since $\mathcal{X} \cong \mathcal{F}_4$ in characteristic 3, its group of automorphisms over $\mathbb{F}_9$ must accommodate both G and the 96-element group of automorphisms of $\mathcal{F}_4$ in characteristic zero. In fact $\mathrm{Aut}_{\mathbb{F}_9}(\mathcal{X})$ is the considerably larger group $\mathrm{U}_3(3)$ of order 6048, consisting of the unitary 3×3 matrices over $\mathbb{F}_9$; it is the largest automorphism group of any genus-3 cuver over an arbitrary field. Again this is a special case of the remarkable behavior of the "Hermitian curve" $\mathcal{F}_{q+1}/\mathbb{F}_{q^2}$: by regarding $x^q x+y^q y+z^q z$ as a ternary Hermitian form over $\mathbb{F}_{q^2}$ we see that any linear transformation of x, y, z which preserves this form up to scalar multiples also preserves the zero-locus (3.19); since of those transformations only multiples of the identity act trivially on $\mathcal{F}_{q+1}$, we conclude that the group $\mathrm{PGU}_3(q)$ acts on $\mathcal{F}_{q+1}$ over $\mathbb{F}_{q^2}$. Once $q>2$, this is the full $\overline{\mathbb{F}}_q$-automorphism group of $\mathcal{F}_{q+1}$, and is the only example of a group of order $>16w^4$ acting on a curve of genus $w>1$ (here $w=(q^2-q)/2$ and the group has order $q^3(q^2-1)(q^3+1)$) over an arbitrary field [Stichtenoth 1973].

Returning to the special case of $\mathcal{X}$, we note that the stabilizer in G of each of its 28 $\mathbb{F}_9$-rational points is a subgroup $N(H_3)\cong S_3$. Thus the two fixed points on $\mathcal{X}$ of H_3 collapse mod 3 to a single point; for each of the three involutions in S_3, this point is also the reduction of one of its four fixed points. Thus the 56- and 84-point G-orbits reduce mod 3 to the same 28-point orbit. The 24-point orbit is undisturbed, and is first seen in $\mathcal{X}(\mathbb{F}_{9^3})$; all other points of $\mathcal{X}$ in characteristic 3 have trivial stabilizer.

Since E_k is supersingular in characteristic 3, its ring of $\bar{\mathbb{F}}_3$-endomorphisms has rank 4 instead of 2; thus the Mordell–Weil lattice of $\bar{\mathbb{F}}_3$-maps from $\mathcal{X}$ to E_k

now has rank 12 instead of 6. Gross [1990, p. 957] used the action of $U_3(3)$ on this lattice to identify it with the Coxeter–Todd lattice. This lattice has $756 = 63 \cdot 12$ minimal vectors of norm 4, which as before come from involutions of the curve; the count is higher than in characteristic zero because there are 63 involutions in $U_3(3) = \mathrm{Aut}_{\bar{\mathbb{F}}_3} \mathcal{X}$ and 12 automorphisms of E_k, rather than 21 and 2 respectively. To see the new automorphisms, reduce (2.10) mod 3 to obtain $y^2 = x^3 + x$, with automorphisms generated by $(x, y) \mapsto (x + 1, y)$ and $(x, y) \mapsto (-x, iy)$ with $i^2 = -1$.

Adler [1997] found that the modular curve $X(11)$, with automorphism group $\mathrm{PSL}_2(\mathbb{F}_{11})$ in characteristic zero, has the larger automorphism group M_{11} when reduced mod 3. Once we identify $\mathcal{X}$ with the modular curve $X(7)$ in the next section we'll be able to regard its extra automorphisms mod 3 as a similar phenomenon. This quartic in characteristic 3 has another notable feature: each of its points is an inflection point! See [Hartshorne 1977, p. 305, Ex. 2.4], where the curve[13] is described as "funny" for this reason. (The 28 points of $\mathcal{X}(\mathbb{F}_9)$ are distinguished in that their tangents meet $\mathcal{X}$ with multiplicity 4 instead of 3; these fourfold tangents are the reductions mod 3 of the bitangents of $\mathcal{X}$ in characteristic zero.) Again Adler found in [1997] that $X(11)$, naturally embedded in the 5-dimensional representation of $\mathrm{PSL}_2(11)$, is also "funny" in this sense when reduced mod 3. While it is not reasonable to expect the extra automorphisms of $X(7)$ and $X(11)$ in characteristic 3 to generalize to higher modular curves $X(N)$ (the Mathieu group M_{11}, being sporadic, can hardly generalize), one might ask whether further modular curves are "funny" mod 3 or in other small characteristics.

Characteristic 7. The curve $\mathcal{X}$ even has good reduction in characteristic 7 over a large enough extension of $\mathbb{Q}$; that is, $\mathcal{X}$ has "potentially good reduction" mod 7. We can see this from our realization of $\mathcal{X}$ as a cyclic triple cover of E_k. The elliptic curve E_k has potentially good reduction mod 7, because the change of variable $x = \sqrt{-7}\, x_1$ puts its Weierstrass equation (2.10) in the form

$$(\sqrt{-7}\,)^{-3} y^2 = 4x_1^3 - 3\sqrt{-7}\, x_1^2 - 4x_1, \tag{3.21}$$

and over a number field containing $(-7)^{1/4}$ the further change of variable $y = 2(-7)^{3/4} y_1$ makes (3.21) reduce to $y_1^2 = x_1^3 - x_1$ at a prime above 7. [In general any CM elliptic curve has potentially good reduction at all primes; equivalently, the j-invariant of any CM curve is an algebraic integer.] Since the x-coordinates of the two branch points of the cover $\mathcal{X} \to E_k$ are the roots of (2.11), their x_1 coordinates are the roots of $\sqrt{-7} x_1^2 + x_1 = \sqrt{-7}$, one of which has negative 7-valuation while the other's 7-valuation is positive. Thus these points reduce to distinct points on $y_1^2 = x_1^3 - x_1$, namely the point at ∞ and the 2-torsion point $(0, 0)$, and the cover $\mathcal{X} \to E_k$ branched at those points has good reduction mod 7 as well.

[13] In its Klein model, but for once the distinction is irrelevant.

On the other hand, the homogeneous quartic defining $\mathcal{X}$ cannot have good reduction mod 7, even potentially: we have seen that even in the rational S_3 model the quartic invariant Φ_4 factors mod 7 as Φ_2^2. How can a plane quartic curve have good reduction if its defining equation becomes so degenerate?

This apparent paradox is resolved only by realizing that the moduli space of curves of genus 3 contains not only plane quartics but also hyperelliptic curves. While a non-hyperelliptic curve of genus 3 is embedded as a quartic in $\mathbb{P}^2$ canonically[14], the canonical map to $\mathbb{P}^2$ from a hyperelliptic curve of genus 3 is a double cover of a conic $C : Q_2 = 0$. Moreover, the moduli space of curves of genus 3 is connected, so a hyperelliptic curve S of genus 3 may be contained in a one-parameter family of curves of the same genus most of which are not hyperelliptic. In that case, the neighbors of S in the family are plane quartics $Q_4 = 0$ that approach the double conic $Q_2^2 = 0$ coming from S; if we write Q_4 as $Q_2^2 + \varepsilon Q_4' + O(\varepsilon^2)$ in a neighborhood of S then the branch points of the double cover $S \to C$ are the $2 \cdot 4 = 8$ zeros of Q_4' on C.[15] This means that a smooth plane quartic curve $Q_4 = 0$ may reduce to a hyperelliptic curve of genus 3 modulo a prime at which $Q_4 \equiv Q_2^2$. This is in fact the case for our curve $\mathcal{X}$, with $Q_4 = \Phi_4$ and $Q_2 = \Phi_2$: Serre found [Mazur 1986, p. 238, footnote] that, over an extension of k sufficiently ramified above $\wp_7$, the Klein quartic reduces to

$$v^2 = u^7 - u \tag{3.22}$$

at that prime, where u is a degree-1 function on the conic $\Phi_2 = 0$ in $\mathbb{P}^2$ that identifies this conic with $\mathbb{P}^1$. This reduced curve (3.22) inherits the action of G: a group element $\pm\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{PSL}_2(\mathbb{F}_7)$ acts on (3.22) by

$$\pm\begin{pmatrix} a & b \\ c & d \end{pmatrix} : (u, v) \mapsto \left(\frac{au+b}{cu+d}, \frac{v}{(cu+d)^4}\right). \tag{3.23}$$

As in the case of characteristic 3, the group of automorphisms of the reduced curve properly contains G; here it is the direct product of G with the two-element group $(u, v) \mapsto (u, \pm v)$ generated by the hyperelliptic involution. Also as in characteristic 3, this reduced curve attains the upper or lower Weil bound on the number of points of a genus-3 curve over finite fields of even degree over the prime field. This is because the prime 7 is not split in k, so the reduction of E_k to an elliptic curve in characteristic 7 is supersingular. The supersingularity could also be seen directly from its Weierstrass model $y_2^2 = x_1^3 - x_1$; that the eigenvalues of Frobenius for $v^2 = u^7 - u$ over $\mathbb{F}_{49}$ all equal -7 could also be seen by counting points: since $(u^7 - u)/\sqrt{-1} \in \mathbb{F}_7$ for all $\mathbb{F}_{49}$, and $\sqrt{-1}$ is a square in $\mathbb{F}_{49}$, the preimages of each $u \in \mathbb{P}^1(\mathbb{F}_{49}) - \mathbb{P}^1(\mathbb{F}_7)$ are $\mathbb{F}_{49}$-rational, and these

[14] Here "canonically" means via curve's holomorphic differentials, which are sections of the canonical divisor; see for instance [Hartshorne 1977, p. 341].

[15] Thanks to Joe Harris for explaining this point; it should be well-known, but is not easy to find in the literature. Armand Brumer points out that this picture is explained in [Clemens 1980, pp. 155–157].

$2 \cdot 42 = 84$ points together with the 8 Weierstrass points $u \in \mathbb{P}^1(\mathbb{F}_7)$ add up to 92, which attains the Weil bound $49 + 6 \cdot 7 + 1$.

4. $\mathcal{X}$ as a Modular Curve

4.1. $\mathcal{X}$ as the modular curve $X(7)$. Since $G \cong \mathrm{PSL}_2(\mathbb{F}_7)$ we can realize G as the quotient group $\Gamma(1)/\Gamma(7)$, where $\Gamma(1)$ is the modular group $\mathrm{PSL}_2(\mathbb{Z})$ and $\Gamma(7)$ is the subgroup of matrices congruent to the identity mod 7. The following facts are well known: $\Gamma(1)$ acts on the upper half-plane $\mathcal{H} = \{\tau \in \mathbb{C} : \operatorname{Im} \tau > 0\}$ by fractional linear transformations $\pm\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) : \tau \mapsto (a\tau + b)/(c\tau + d)$; the quotient curve $\mathcal{H}/\Gamma(1)$ parametrizes elliptic curves up to $\mathbb{C}$-automorphism; if we extend $\mathcal{H}$ by to $\mathcal{H}^*$ by including the "cusps" $\mathbb{Q} \cup \{\infty\} = \mathbb{P}^1(\mathbb{Q})$, the resulting quotient curve $X(1)$ may be regarded as a compact Riemann surface of genus 0; and for each $N \geq 1$, the quotient of $\mathcal{H}^*$ by the normal subgroup $\Gamma(N)$ of $\Gamma(1)$ is the modular curve $X(N)$ whose non-cusp points parametrize elliptic curves E with a full level-N structure. A "full level-N structure" means an identification of the group $E[N]$ of N-torsion points with some fixed group T_N. Why T_N and not simply $(\mathbb{Z}/N)^2$ as expected? We can certainly use $T_N = (\mathbb{Z}/N)^2$ if we regard $X(N)$ as a curve over an algebraically closed field such as $\mathbb{C}$. But that will not do over $\mathbb{Q}$ once $N > 2$: the *Weil pairing* (see for instance [Silverman 1986, III.8, pp. 95 ff.]) identifies $\wedge^2 E[N]$ with the N-th roots of unity $\boldsymbol{\mu}_N$, which are not contained in $\mathbb{Q}$. So T_N must be some group $\cong (\mathbb{Z}/N)^2$ equipped with an action of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ such that $\wedge^2 T_N \cong \boldsymbol{\mu}_N$ as Galois modules. There are many choices for T_N — for instance, $E[N]$ for any elliptic curve $E/\mathbb{Q}$! — which in general yield different modular curves over $\mathbb{Q}$ (though they all become isomorphic over $\overline{\mathbb{Q}}$): T_N and T'_N yield the same curve only if $T'_N \cong T_N \otimes \psi$ for some quadratic character ψ. The simplest choice is

$$T_N = (\mathbb{Z}/N) \times \boldsymbol{\mu}_N, \tag{4.1}$$

and that is the choice that we shall use to define X_N as a curve over $\mathbb{Q}$. Note, however, that the action of $\Gamma(1)/\Gamma(N)$ is still defined only over the cyclotomic field $\mathbb{Q}(\boldsymbol{\mu}_N)$. The canonical map $X(N) \to X(1)$ that forgets the level-N structure is a Galois cover with group $\Gamma(1)/\Gamma(N) = \mathrm{PSL}_2(\mathbb{Z}/N)$; it is ramified only above three points of $X(1)$, namely the cusp and the elliptic points that parametrize elliptic curves with complex multiplication by $\mathbb{Z}[i]$ and $\mathbb{Z}[e^{2\pi i/3}]$, and the ramification indices at these points are N, 2, and 3 respectively.

Now consider $N = 7$. Then $X(7)$ is a G-cover of the genus-0 curve $X(1)$ with three branch points of indices $2, 3, 7$; therefore it is a Hurwitz curve, and thus isomorphic with $\mathcal{X}$ at least over $\mathbb{C}$. The 24-point orbit is the preimage of the cusp, and the 56- and 84-point orbits are the preimages of the elliptic points $\tau = e^{2\pi i/3}$ and $\tau = i$ on $X(1)$ parametrizing CM elliptic curves with j-invariants 0 and 1728. We shall show that the choice (4.1) of T_7 yields $X(7)$ as a curve over $\mathbb{Q}$ isomorphic with the Klein model of $\mathcal{X}$, and give explicitly an elliptic curve and 7-torsion points parametrized by a generic point $(x:y:z) \in \mathbb{P}^2$ with $x^3y + y^3z + z^3x = 0$.

The projective coordinates for $\mathcal{X}$ can be considered as a basis for $H_1(\mathcal{X})$. Holomorphic differentials on a modular curve $\mathcal{H}^*/\Gamma$ are differentials $f(\tau)\,d\tau$ on $\mathcal{H}^*$ that are regular and invariant under Γ, i.e. such that $f(\tau)$ is a *modular cusp form of weight* 2 for Γ: a holomorphic function satisfying the identity

$$f\left(\frac{a\tau+b}{c\tau+d}\right) = (c\tau+d)^2 f(\tau) \tag{4.2}$$

for all $\pm\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ and vanishing at the cusps. We next choose a convenient basis for the modular cusp forms of weight 2 for $\Gamma(7)$.

Taking $\pm\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \pm\left(\begin{smallmatrix} 1 & 7 \\ 0 & 1 \end{smallmatrix}\right)$ in (4.2) we see that f must be invariant under $\tau \mapsto \tau + 7$; thus it has a Fourier expansion in powers of $q^{1/7}$, where as usual

$$q := e^{2\pi i\tau} \qquad (\text{so} \quad d\tau = \frac{1}{2\pi i}\frac{dq}{q}). \tag{4.3}$$

Since we require vanishing at the cusp $\tau = i\infty$, the expansion must involve only positive powers of $q^{1/7}$. The action of $g = \pm\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ on modular forms multiplies q by ζ; thus g decomposes our space of modular forms into eigen-subspaces with eigenvalues ζ^a, such that for each $a \bmod 7$ the ζ^a eigenspace consists of forms $\sum_{m>0} c_m q^{m/7}$ whose coefficients c_m vanish at all $m \not\equiv a \bmod 7$. We find three such forms:

$$\begin{aligned}
\mathsf{x} &= q^{4/7}(-1 + 4q - 3q^2 - 5q^3 + 5q^4 + 8q^6 - 10q^7 + 4q^9 - 6q^{10} \cdots),\\
\mathsf{y} &= q^{2/7}(1 - 3q - q^2 + 8q^3 - 6q^5 - 4q^6 + 2q^8 + 9q^{10} \cdots),\\
\mathsf{z} &= q^{1/7}(1 - 3q + 4q^3 + 2q^4 + 3q^5 - 12q^6 - 5q^7 + 7q^9 + 16q^{10} \cdots).
\end{aligned} \tag{4.4}$$

These can be expressed as the modified theta series

$$\mathsf{x}, \mathsf{y}, \mathsf{z} = \sum_{\beta} \operatorname{Re}(\beta)\, q^{\beta\bar{\beta}/7}, \tag{4.5}$$

the sum extending over $\beta \in \mathbb{Z}[\alpha]$ congruent mod $(\sqrt{-7})$ to $2, 4, 1$ respectively. They also have the product expansions

$$\mathsf{x}, \mathsf{y}, \mathsf{z} = \varepsilon q^{a/7} \prod_{n=1}^{\infty} (1-q^n)^3 (1-q^{7n}) \prod_{\substack{n>0 \\ n \equiv \pm n_0 \bmod 7}} (1-q^n), \tag{4.6}$$

where the parameters ε, a, n_0 are: for x, $-1, 4, 1$; for y, $+1, 2, 2$; and for z, $+1, 1, 4$. That these in fact yield modular forms can be seen by factoring the resulting products (4.6) into Klein forms (for which see for instance [Kubert and Lang 1981, pp. 25 ff. and 68 ff.]); it follows that $\mathsf{x}, \mathsf{y}, \mathsf{z}$ do not vanish except at the cusps of $X(7)$.

Since $\mathsf{x}, \mathsf{y}, \mathsf{z}$ are ζ^4-, ζ^2-, and ζ-eigenforms for g, the three-dimensional representation of G that they generate must be isomorphic with (V, ρ), and so the action of G on $X(7)$ will make it a quartic in the projectivization not of (V, ρ) but of (V^*, ρ^*).[16] Using either the sum or the product formulas for $\mathsf{x}, \mathsf{y}, \mathsf{z}$, together with the action of $\Gamma(1)$ on theta series or on Klein forms, we can compute that h cyclically permutes $\mathsf{x}, \mathsf{y}, \mathsf{z}$. This is enough to identify $(\mathsf{x}, \mathsf{y}, \mathsf{z})$ up to scaling with our standard basis for V (again thanks to the fact that the 21-element subgroup $\langle g, h\rangle$ of G acts irreducibly on V). This leads us to expect that

$$\Phi_4(\mathsf{x}, \mathsf{y}, \mathsf{z}) = \mathsf{x}^3\mathsf{y} + \mathsf{y}^3\mathsf{z} + \mathsf{z}^3\mathsf{x} = 0, \tag{4.7}$$

and the q-expansions corroborate this. To prove it we note that $\Phi_4(\mathsf{x}, \mathsf{y}, \mathsf{z})$, being a G-invariant polynomial in the cusp forms $\mathsf{x}, \mathsf{y}, \mathsf{z}$, must be a cusp form of weight $4 \cdot 2 = 8$ for the full modular group $\Gamma(1)$; but the only such form is zero. (See for instance [Serre 1973, Ch.VII] for the complete description of cusp forms on $\Gamma(1)$.) Thus the coordinates $(\mathsf{x} : \mathsf{y} : \mathsf{z})$ for the canonical image of $X(7)$ in $\mathbb{CP}^2$ identify it with the Klein model of $\mathcal{X}$.

We next identify the other G-invariant polynomials in $\mathsf{x}, \mathsf{y}, \mathsf{z}$ with known modular cusp forms for $\Gamma(1)$. We find that

$$\Phi_6(\mathsf{x}, \mathsf{y}, \mathsf{z}) = \Delta\ [= q\prod_{n=1}^{\infty}(1-q^n)^{24} = q - 24q^2 + 252q^3 \cdots], \tag{4.8}$$

which requires only checking the q^1 coefficient because every $\Gamma(1)$ cusp form of weight 12 is a multiple of Δ. Likewise the leading terms of $\Phi_{14}(\mathsf{x}, \mathsf{y}, \mathsf{z})$ and $\Phi_{21}(\mathsf{x}, \mathsf{y}, \mathsf{z})$, together with their weights 28, 42, suffice to identify these modular forms with

$$\Phi_{14}(\mathsf{x}, \mathsf{y}, \mathsf{z}) = \Delta^2 E_2\ [= \Delta^2\Big(1 + 240\sum_{n=1}^{\infty}\frac{n^3q^n}{1-q^n}\Big) = q^2 + 192q^3 - 8280q^4 \cdots], \tag{4.9}$$

$$\Phi_{21}(\mathsf{x}, \mathsf{y}, \mathsf{z}) = \Delta^3 E_3\ [= \Delta^3\Big(1 - 504\sum_{n=1}^{\infty}\frac{n^5q^n}{1-q^n}\Big) = q^3 - 576q^4 + 22140q^5 \cdots]. \tag{4.10}$$

Thus the elliptic curve parametrized by a non-cusp point $(x : y : z)$ on $\mathcal{X}$ is

$$E_{(x:y:z)} :\ v^2 = u^3 - \tfrac{1}{48}\lambda^2\Phi_{14}(x, y, z) + \tfrac{1}{864}\lambda^3\Phi_{21}(x, y, z), \tag{4.11}$$

for some yet unknown λ of weight -14 (that is, homogeneous of degree -7 in x, y, z) that only changes $E_{(x:y:z)}$ by a quadratic twist.

To determine the values of u at 7-torsion points of $E_{(x:y:z)}$ we identify that curve with $\mathbb{C}/(\mathbb{Z} \oplus \mathbb{Z}\tau) \cong \mathbb{C}^*/q^{\mathbb{Z}}$ and expand the Weierstrass $\wp$-function of that curve at some point $q_1 \in \mathbb{C}^*/q^{\mathbb{Z}}$ in a q-series depending on q_1. We find

$$u = \lambda\Delta\bigg(\frac{1}{12} - 2\sum_{n=1}^{\infty}\frac{q^n}{(1-q^n)^2} + \sum_{n=-\infty}^{\infty}\frac{q^nq_1}{(1-q^nq_1)^2}\bigg). \tag{4.12}$$

[16]This mildly unfortunate circumstance could only have been avoided by more awkward artifices such as declaring ζ to be $e^{-2\pi i/7}$ instead of $e^{+2\pi i/7}$ in (0.1). Of course the distinction between the V and V^* models of $\mathcal{X}$ is harmless because the two representations are related by an outer automorphism of G.

The 7-torsion points of $\mathbb{C}^*/q^{\mathbb{Z}}$ are generated by ζ and $q^{1/7}$. Substituting these for q_1 in (4.12) we obtain $\lambda P(\mathsf{x},\mathsf{y},\mathsf{z})$ for certain polynomials P of degree 7 determined up to multiples of Φ_4. We find P by comparing q-expansions. For $q_1 = \zeta$ we obtain the symmetrical form

$$P = \tfrac{1}{7}\big((c_1 - 2c_2 - \tfrac{53}{12})\mathsf{x}^7 + (c_2 - 2c_4 - \tfrac{53}{12})\mathsf{y}^7 + (c_4 - 2c_1 - \tfrac{53}{12})\mathsf{z}^7\big) \\ + \tfrac{2}{3}\big((c_2 - c_4)\mathsf{x}^4\mathsf{y}^2\mathsf{z} + (c_4 - c_1)\mathsf{y}^4\mathsf{z}^2\mathsf{x} + (c_1 - c_2)\mathsf{z}^4\mathsf{x}^2\mathsf{y}\big), \quad (4.13)$$

using the abbreviation $c_j := \zeta^j + \zeta^{-j} \in K_+$. the polynomials for ζ^2, ζ^4 are obtained from these by cyclically permuting c_1, c_2, c_4 and $\mathsf{x},\mathsf{y},\mathsf{z}$. That only these six monomials can occur is forced by the invariance of the polynomial under $\langle g \rangle$. The polynomial for $q_1 = q^{1/7}$ looks more complicated, because invariance under sgs is not so readily detectable; we refrain from exhibiting that polynomial in full, but note that it can be obtained from (4.13) by the linear substitution $\rho(s)$, and that its coefficients, unlike those of (4.13), are *rational.*[17]

It remains to choose λ. We would have liked to make it G-invariant, since the action of G would then preserve our model (4.11) for $E_{(x:y:z)}$ and only permute its 7-torsion points. But we cannot make λ an arbitrary homogeneous function of degree -7 in $\mathsf{x},\mathsf{y},\mathsf{z}$ because we are constrained by the condition that $E_{(x:y:z)}[7] \cong T_7$ for all non-cusp $(x : y : z) \in X(7)$. This means, first, that $E_{(x:y:z)}$ must be a nondegenerate elliptic curve, and second, that its 7-torsion group be generated by a rational point (for the $\mathbb{Z}/7$ part of T_7) and a point that every $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ element taking ζ to ζ^a multiplies by a (for the $\boldsymbol{\mu}_7$ part). The first condition amounts to the requirement that the divisor of λ be supported on the cusps of $X(7)$; this determines λ up to multiplication by a "modular unit" in $\mathbb{C}(X(7))$. The second condition then determines λ up to multiplication by the square of a modular unit. It turns out that already the first condition prevents us from choosing a G-invariant λ: such a λ would be Φ_{14}/Φ_{21} times a rational function of j, and thus would have zeros or poles on the elliptic points of order 2 and 3 (the 56- and 84-point orbits of $\mathcal{X}$).

We next find a λ, necessarily not G-invariant, that does the job. From our computation of u-coordinates at 7-torsion points we know that $u/\lambda\Delta$ is a polynomial $P(\mathsf{x},\mathsf{y},\mathsf{z}) \in \mathbb{Q}[\mathsf{x},\mathsf{y},\mathsf{z}]$. Moreover

$$Q := P^3 - \tfrac{1}{48}\Phi_{14}P + \tfrac{1}{864}\Phi_{21} \quad (4.14)$$

cannot vanish except at a cusp, lest a 7- and a 2-torsion point on $\mathbb{C}^*/q^{\mathbb{Z}}$ coincide. [In fact Q has the product expansion

$$q^{23/7}\prod_{n=1}^{\infty}\big((1-q^{n-\frac{6}{7}})(1-q^{n-\frac{1}{7}})\big)^{-8}\big((1-q^{n-\frac{5}{7}})^2(1-q^{n-\frac{2}{7}})\big)^2(1-q^n)^{84}, \quad (4.15)$$

[17] This is ultimately due to the fact that the coefficients of (4.12) are rational. In fact it is no accident the least common denominator of the coefficients of P for $q_1 = q^{1/7}$ is 12, same as for (4.12); but we need not pursue this here.

which manifestly has neither zero nor pole in $X(7)-\{\text{cusps}\}$.] Thus for any λ_0 homogeneous of degree 14 in $\mathsf{x}, \mathsf{y}, \mathsf{z}$ whose divisor is supported on the cusps (for instance $\lambda_0 = \mathsf{x}^{14}$) we may take

$$\lambda = Q/\lambda_0^2, \tag{4.16}$$

which satisfies the first condition and yields a 7-torsion point on the curve (4.11) rational over $\mathbb{Q}(x, y, z)$.

We claim that this, together with our computations thus far, lets us deduce that λ also satisfies the second condition, and thus completes our proof that $X(7)$ is $\mathbb{Q}$-isomorphic with $\mathcal{X}$, as well as the determination of the 7-torsion points on the generic elliptic curve (4.11) parametrized by $\mathcal{X}$. We must show that $E[7]$ is isomorphic as a $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ module with $T_7 = (\mathbb{Z}/7) \times \boldsymbol{\mu}_7$. Indeed, consider the action on $E[7]$ of an element γ of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ that takes ζ to ζ^a. By our choice of λ, this γ fixes the point with $q_1 = q^{1/7}$; thus this point generates a subgroup $\cong \mathbb{Z}/7$ of $E[7]$. From our computation of (4.13) we see that γ multiplies the $q_1 = \zeta$ point by either a or $-a$. But the Weil pairing of the ζ and $q^{1/7}$ points is ζ, which γ takes to ζ^a. Thus γ must also take the $q_1 = \zeta$ point to ζ^a. In other words, the $q_1 \in \boldsymbol{\mu}_7$ points comprise a subgroup of $E[7]$ isomorphic as a Galois module with $\boldsymbol{\mu}_7$. Having found subgroups of $E[7]$ isomorphic with $\mathbb{Z}/7$ and $\boldsymbol{\mu}_7$, we are done.

4.2. The modular interpretation of quotients of $\mathcal{X}$. Now let H be a subgroup of G, and consider the quotient curve $\mathcal{X}/H$. When H is trivial, this quotient is $\mathcal{X}$ itself, which we have just identified with the moduli space $X(7)$ of elliptic curves with full level-7 structure. When $H = G$, the quotient is the moduli space $X(1)$ of elliptic curves with no further structure, and the quotient map $X(7) \to X(1)$ in effect forgets the level-7 structure. For intermediate groups H, the quotient curve, which can still be regarded also as the quotient of $\mathcal{H}^*$ by a congruence subgroup of $\Gamma(1)$, parametrizes elliptic curves with partial level-7 structure such as a choice of a 7-torsion point or 7-element subgroup. In this section we describe the three classical modular curves $X_0(7)$, $X_1(7)$, and $X_0(49)$ that arise in this way. The same constructions yield for each $N > 1$ the curves $X_0(N)$, $X_1(N)$, $X_0(N^2)$ as quotients of $X(N)$, though of course for each N we face anew the problem of finding explicit coordinates and equations for these modular curves and covers.

Each of the eight 7-element subgroups T of E (equivalently, of $E[7]$) yields an isogeny of degree 7 from E to the quotient elliptic curve E/T. The T's may be regarded as points of the projective line $(E[7] - \{\mathbf{0}\})/\mathbb{F}_7^* \cong \mathbb{P}^1(\mathbb{F}_7)$, permuted by G. The stabilizer in G of a point on this $\mathbb{P}^1(\mathbb{F}_7)$ is a 21-element subgroup; for instance, $\langle g, h \rangle$ is the stabilizer of ∞. Taking $H = \langle g, h \rangle$ we conclude that $\mathcal{X}/H$ parametrizes elliptic curves E together with a 7-element subgroup T, or equivalently together with a 7-isogeny $E \to E/T$. This $\mathcal{X}/H$ is the quotient of $\mathcal{H}^*$

by the subgroup

$$\Gamma_0(7) := \left\{ \pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{Z}) : c \equiv 0 \bmod 7 \right\} \tag{4.17}$$

of $\Gamma(1)$, and is called the modular curve $X_0(7)$. This curve has genus 0, with rational coordinate ("Hauptmodul")

$$j_7 = \frac{1}{q}\left(\prod_{n=1}^{\infty}(1-q^n)/(1-q^{7n})\right)^4 = q^{-1} - 4 + 2q + 8q^2 - 5q^3 - 4q^4 \cdots. \tag{4.18}$$

Comparing this with the product expansions for $\mathsf{x}, \mathsf{y}, \mathsf{z}, \Delta$, we may express j_7 as a quotient of $\langle g, h\rangle$-invariant sextics in $\mathsf{x}, \mathsf{y}, \mathsf{z}$:

$$j_7 = \frac{(\mathsf{xyz})^2}{\Delta} = \frac{(\mathsf{xyz})^2}{\Phi_6(\mathsf{x}, \mathsf{y}, \mathsf{z})}. \tag{4.19}$$

Either by comparing this with (2.13), or directly from the q-expansions, we then find that the degree-8 cover $X_0(7)/X(1)$ is given by

$$j = (j_7^2 + 13j_7 + 49)(j_7^2 + 245j_7 + 7^4)^3/j_7^7. \tag{4.20}$$

Given a 7-isogeny $E \to E/T$, the image of $E[7]$ in E/T is a 7-element subgroup of E/T and thus yields a new 7-isogeny $E/T \to E/E[7] \cong E$. This is in fact the *dual isogeny* [Silverman 1986, p. 84 ff.] of the isogeny $E \to E/T$. Thus we have a rational map $w_7 : X_0(7) \to X_0(7)$ that takes a non-cusp point of $X_0(7)$, parametrizing an isogeny $E \to E/T$, to the point parametrizing the dual isogeny $E/T \to E$. Moreover, iterating this construction recovers our original isogeny $E \to E/T$; thus w_7 is an involution of $X_0(7)$. This w_7 is known as the *Fricke involution* of $X_0(7)$. In general $X_0(N) = \mathcal{H}^*/\Gamma_0(N)$ parametrizes N-isogenies with cyclic kernel (a.k.a. "cyclic N-isogenies") between elliptic curves, and the dual isogeny yields the Fricke involution w_N of $\mathcal{X}_0(N)$. This involution can also be described over $\mathbb{C}$ as the action of the fractional linear transformation $\tau \leftrightarrow -1/N\tau$ on $\mathcal{H}^*$, which descends to an automorphism of $X_0(N)$ because it normalizes $\Gamma_0(N)$. In our case of $N = 7$ we find the formula

$$w_7(j_7) = 49/j_7 \tag{4.21}$$

for the action of w_7 on $X_0(7)$. The coefficients of the curve E/T and the 7-isogenies $E \rightleftharpoons E/T$ parametrized by $X_0(7)$ can be computed as explicit functions of j_7 by the methods of [Elkies 1998a].

The modular curve $X_1(7)$ parametrizes elliptic curves with a rational 7-torsion point. It is thus the quotient of $X(7)$ by the subgroup of G that fixes a 7-torsion point. To obtain this modular curve, and the elliptic curve it parametrizes, over $\mathbb{Q}$, we must be careful to use a 7-torsion point that generates the subgroup $\mathbb{Z}/7$ of T_7: we have already computed in (2.2) the quotient of $\mathcal{X}$ by the 7-element subgroup $\langle g\rangle$ of G, which is the stabilizer of a 7-torsion point; but this is the point (4.13), which generates the subgroup $\boldsymbol{\mu}_7$ of T_7, and so is not rational over $\mathbb{Q}$.

The $\mathbb{Z}/7$ subgroup has stabilizer $\langle sgs \rangle$, so we may obtain $X_1(7)$ as $X(7)/\langle sgs \rangle$. Alternatively we may start from $X(7)/\langle g \rangle$ and apply w_7. This second approach requires some explanation. At the level of Riemann surfaces, there is no problem: for any $N > 1$, the modular curve $X_1(N)$ is $\mathcal{H}^*/\Gamma_1(N)$ where

$$\Gamma_1(N) := \left\{ \pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{Z}) : c \equiv 0,\ a, d \equiv 1 \bmod N \right\}, \tag{4.22}$$

and again $\tau \leftrightarrow -1/N\tau$ normalizes this subgroup and so yields an involution of $X_1(N)$. But over $\mathbb{Q}$ some care is required. The curve $X_1(N)$ parametrizes pairs (E, P) where E is an elliptic curve and $P \in E$ is a point of order N. The involution takes (E, P) to (E', P'), where $E' = E/\langle P \rangle$ and P' generates the image of $E[N]$ under the quotient map $E \to E'$. But to specify the generator we must use the Weil pairing: P' must be the image of a point $\tilde{P} \in E[N]$ whose Weil pairing with P is $e^{2\pi i/N}$. Once $N > 2$ the root of unity $e^{2\pi i/N}$ is not rational, so we cannot demand that both P and P' be rational N-torsion points on E, E'. Instead, P, P' must generate Galois modules such that $\langle P \rangle \otimes \langle P' \rangle \cong \boldsymbol{\mu}_N$. So, for instance, if P is rational then $\langle P' \rangle \cong \boldsymbol{\mu}_N$, and conversely if $\langle P \rangle \cong \boldsymbol{\mu}_N$ then P' is rational. The latter case applies for us: in our model of $X(7)$, the distinguished 7-torsion points on the elliptic curve E parametrized by $X(7)/\langle g \rangle$ constitute a subgroup $\cong \boldsymbol{\mu}_7$ of $E[7]$; thus the curve E' has a rational 7-torsion point.

Using $\mathcal{X}/\langle g \rangle$ for $X_1(7)$, we find that this modular curve has rational coordinate

$$d := \frac{\mathsf{y}^2\mathsf{z}}{\mathsf{x}^3} = q^{-1} + 3 + 4q + 3q^2 - 5q^4 - 7q^5 - 2q^6 + 8q^7 \cdots, \tag{4.23}$$

and that the cyclic cubic cover $X_1(7) \to X_0(7)$ is given by

$$j_7 = d + \frac{1}{1-d} + \frac{d-1}{d} - 8 = \frac{d^3 - 8d^2 + 5d + 1}{d^2 - d}. \tag{4.24}$$

The elliptic curve with a 7-torsion point parametrized by $X_1(7)$ was already exhibited in extended Weierstrass form by Tate [1974, p. 195]:

$$y^2 + (1 + d - d^2)xy + (d^2 - d^3)y = x^3 + (d^2 - d^3)x^2 \tag{4.25}$$

(we chose our coordinate d so as to agree with this formula). Besides making the coefficients simpler compared to the standard Weierstrass form $y^2 = x^3 + a_4x + a_6$, Tate's formula has the advantage of putting the origin at a 7-torsion point — Tate actually obtained (4.25) starting from a generic elliptic curve

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 \tag{4.26}$$

tangent to the x-axis at the origin, and working out the condition for the origin to be a 7-torsion point. The equations for the curve 7-isogenous with (4.25) can again be obtained by the methods of [Elkies 1998a], or — since here the points of the isogeny's kernel are rational — already from Vélu's formulas [Vélu 1971] on which those methods are based.

From our discussion in the previous paragraph, the involution w_7 of $X_1(7)$ cannot be defined over $\mathbb{Q}$, only over K_+. (The full cyclotomic field K is not needed because $X_1(7)$ cannot distinguish a 7-torsion point from its inverse, so only the squares in $(\mathbb{Z}/7)^* = \mathrm{Gal}(K/\mathbb{Q})$ are needed, and they comprise $\mathrm{Gal}(K_+/\mathbb{Q})$; in general for each prime $p \equiv 3 \bmod 4$ the Fricke involution w_p of $X_1(p)$ is defined over the real subfield of the cyclotomic field $\mathbb{Q}(e^{2\pi i/p})$.) In fact there are three choices of w_7, cyclically permuted by $\Gamma_0(7)/\Gamma_1(7)$ (and $\mathrm{Gal}(K_+/\mathbb{Q})$); we calculate that the choice associated with $\tau \leftrightarrow -1/7\tau$ gives

$$w_7(d) = \frac{(4+3c_1+c_2)d - (3+3c_1+c_2)}{d-(4+3c_1+c_2)}, \tag{4.27}$$

where $c_j := \zeta^j + \zeta^{-j} \in K_+$ as in (4.13).

We have seen already that $\mathcal{X}/\langle h\rangle$ coincides with $X_0(49)$, and hinted that this is in fact no mere coincidence. We can now explain this: where a point on $X(7)$ specifies an elliptic curve E together with a basis $\pm\{P_1, P_2\}$ for $E[7]$, the $\langle h\rangle$-orbit of the point specifies only the two subgroups $\langle P_1\rangle$ and $\langle P_2\rangle$ generated by the basis elements. Equivalently, it specifies two elliptic curves $E_1 = E/\langle P_1\rangle$, $E_2/\langle P_2\rangle$ among the eight curves 7-isogenous with E. (Note that $\langle h\rangle$ is the stabilizer in $\mathrm{PSL}_2(\mathbb{F}_7)$ of the two points $0, \infty$ on $\mathbb{P}^1(\mathbb{F}_7)$.) But then we obtain a cyclic 49-isogeny $E_1 \to E_2$ by composing the isogenies $E_1 \to E$, $E \to E_2$. Conversely, any cyclic 49-isogeny between elliptic curves factors as a product of two 7-isogenies and thus comes from a point $\mathcal{X}/\langle h\rangle$. Thus $\mathcal{X}/\langle h\rangle$ is indeed the modular curve $X_0(49)$ parametrizing cyclic 49-isogenies. In this description of $X_0(49)$, the involution w_{49} of $\mathcal{X}/\langle h\rangle$ is the involution we have already constructed from the normalizer of $\langle h\rangle$ in G. Note that w_{49} switches the roles of E_1, E_2 but preserves E. In terms of congruence subgroups of $\Gamma(1)$, the identification of $\mathcal{X}/\langle h\rangle$ with $X_0(49)$ is explained by noting that the congruence groups $\left\{\pm\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in \mathrm{PSL}_2(\mathbb{Z}) : b, c \equiv 0 \bmod 7\right\}$ and $\Gamma_0(49)$ are conjugate in $\mathrm{PSL}_2(\mathbb{R})$ by $\pm 7^{-1/2}\left(\begin{smallmatrix} 7 & 0 \\ 0 & 1\end{smallmatrix}\right) : \tau \mapsto 7\tau$.

Some final remarks on this curve $E_k = \mathcal{X}/\langle h\rangle = X_0(49)$: recall that we showed that its only $\mathbb{Q}$-rational points are the point at infinity and $(0,0)$. Since these are both cusps of $X_0(49)$ we conclude that there are no elliptic curves over $\mathbb{Q}$ admitting a rational cyclic 49-isogeny. However, there are infinitely many number fields, including quadratic ones such as $\mathbb{Q}(i)$ and $\mathbb{Q}(e^{2\pi i/3}) = \mathbb{Q}(\sqrt{-3}\,)$, over which E_k is an elliptic curve of positive rank. (Take $x = -2$ or $x = -3$ in the Weierstrass equation (2.10) for E_k.) Over such a number field there are infinitely many pairs of elliptic curves with different j-invariants that admit a rational cyclic 49-isogeny. Moreover 49 is the largest integer for which this can happen: the curve $X_0(N)$ for $N > 49$ has genus > 1, and thus by Faltings only finitely many points over any given number field. See the tables and introductory remarks of [Birch and Kuyk 1975] for more information on the genera and rational points of the modular curves $X_0(N)$.

4.3. Kenku's proof of the solution of the class number 1 problem. What of the quotients of $\mathcal{X}$ by S_4 and the 2-Sylow subgroup of G? The first of these we calculate using the fact that $\pm\rho(S_4)$ is itself a reflection group, with invariant ring generated by polynomials of degrees $2,4,6$; we choose the elementary symmetric functions of X^2, Y^2, Z^2 as our generators:

$$\Psi_2 := X^2+Y^2+Z^2, \quad \Psi_4 := (XY)^2+(XZ)^2+(YZ)^2, \quad \Psi_6 := (XYZ)^2. \tag{4.28}$$

We then express a basis for the G-invariants in the S_4 model as polynomials in Ψ_2, Ψ_4, Ψ_6. Clearly the invariant quartic (1.11) is $\Psi_2^2 + (3\alpha-2)\Psi_4$. The degree-6 invariant is proportional to $(1+\alpha)\Psi_2^3 + (2-3\alpha)\Psi_2\Psi_4 - (42+7\alpha)\Psi_6$. The determinant (1.14) defining the degree-14 invariant is proportional to

$$\begin{aligned}(-9+9\alpha)\Psi_2^7 + (56-70\alpha)\Psi_2^5\Psi_4 - (294+105\alpha)\Psi_2^3\Psi_4^2 + (28+154\alpha)\Psi_2\Psi_4^3 \\ +\Psi_6\big((1008+2198\alpha)\Psi_2^4 + (1148-7014\alpha)\Psi_2^2\Psi_4 + (-12348+1078\alpha)\Psi_4^2\big) \\ +(15778+15435\alpha)\Psi_2\Psi_6^2.\end{aligned} \tag{4.29}$$

Now the genus-0 curve $\mathcal{X}/S_4$ is rationally parametrized by the function $f := \Psi_2^3/\Psi_6$, which is of degree 24 on $\mathcal{X}$ and thus of degree 1 on $\mathcal{X}/S_4$. So to obtain the degree-7 cover $\mathcal{X}/S_4 \to \mathcal{X}/G$ we need only write the rational parameter Φ_{14}^3/Φ_6^7 of $\mathcal{X}/G$ as a rational function of Ψ_2^3/Ψ_6 on $\mathcal{X}$. Since $\Psi_2^2 = (2-3\alpha)\Psi_4$ on $\mathcal{X}$, our expressions for the G-invariant polynomials of degrees $6, 14$ simplify to multiples of

$$\Psi_2^3(1+(-14+7\alpha)f), \quad \Psi_2^7\big(3+(490+196\alpha)f+(3430+2401\alpha)f^2\big). \tag{4.30}$$

Thus $j = \Phi_{14}^3/\Phi_6^7$ is given by

$$2^6\big(3+(490+196\alpha)f+(3430+2401\alpha)f^2\big)^3\big/\big(1+(-14+7\alpha)f\big)^7, \tag{4.31}$$

in which the coefficient 2^6 may either be obtained by keeping track of all the constants of proportionality along the way, or by requiring that the third point of ramification of j (other than the points $j=0,\infty$ forced by the factorization in (4.18)) occur at $j=12^3$. To put (4.31) in a nicer form we replace f by the equivalent coordinate ψ, related with f by

$$f = \frac{(\alpha+3)\psi+14+26\alpha}{56(\psi+3(1+\alpha))}, \tag{4.32}$$

which puts the pole of j at $\psi=\infty$ and thus makes j a seventh-degree polynomial in ψ:

$$\begin{aligned} j &= \big(\psi-3(1+\alpha)\big)\big(\psi-(2+\alpha)\big)^3\big(\psi+(3+2\alpha)\big)^3 \\ &= 12^3 + \big(\psi+(2+4\alpha)\big)\big(\psi^2+2\alpha\psi-(6+9\alpha)\big)\big(\psi^2-2(1+\alpha)\psi+(1-2\alpha)\big)^2. \end{aligned} \tag{4.33}$$

We noted already that the S_4 model of $\mathcal{X}$ cannot be defined over $\mathbb{Q}$ because S_4 is its own normalizer in $\mathrm{Aut}(G)$. For the same reason this polynomial (4.33) cannot have rational coefficients. Over a number field F containing k, we may choose a conjugacy class of subgroups $S_4 \subset G$, and then depending on our choice either

(4.33) or its $\mathrm{Gal}(k/\mathbb{Q})$ conjugate parametrizes elliptic curves E/F such that $\mathrm{Gal}(\bar{F})/F$ acts on $E[7]$ by a subgroup of a 24-element group in that conjugacy class.[18]

On the other hand, the 8-element dihedral subgroups D_8 of G do extend to 16-element subgroups of $\mathrm{Aut}(G)$. This is a consequence of Sylow theory, but the subgroups in question can also be seen from the interpretation of G and $\mathrm{Aut}(G)$ as $\mathrm{PSL}_2(\mathbb{F}_7), \mathrm{PGL}_2(\mathbb{F}_7)$: choose an identification of $\mathbb{F}_7^2$ with $\mathbb{F}_{49}$, and consider the action of $\Gamma\mathrm{L}_1(\mathbb{F}_{49})$ on $\mathbb{F}_{49}$. Multiplication by some $a \in \mathbb{F}_{49}^*$ and Galois conjugation are $\mathbb{F}_7$-linear transformations of determinant a^8 and -1 respectively. Using only $\mathbb{F}_{49}^*$ we obtain cyclic subgroups of orders 4, 8 in $\mathrm{PSL}_2(\mathbb{F}_7)$ and $\mathrm{PGL}_2(\mathbb{F}_7)$, the *nonsplit Cartan subgroups* of these linear groups; allowing also Galois conjugation, we obtain the normalizers of the nonsplit Cartan subgroups, which are 8- and 16-element dihedral groups and are the 2-Sylow subgroups of $\mathrm{PSL}_2(\mathbb{F}_7)$, $\mathrm{PGL}_2(\mathbb{F}_7)$ respectively. Since $D_8 \subset G$ is normalized by outer automorphisms of G, the quotient of $\mathcal{X}/D_8$ can be defined over $\mathbb{Q}$ — even though it factors through the quotient by S_4, which is only defined over k! To obtain that quotient as a degree-3 cover of the ψ-line we may either proceed as we did to obtain (4.33), namely, writing Ψ_2, Ψ_4, Ψ_6 in terms of the invariants of D_8, or locate the ramification points of the cover. This triple cover is totally ramified at the simple root $\psi = 3(1+\alpha)$ of j, and has double points at the solutions of $\psi^2 + 2\alpha\psi = (6+9\alpha)$ at which $j = 12^3$. We find that the cover is given by

$$\psi = \frac{(2+3\alpha)\phi^3 - (18+15\alpha)\phi^2 + (42+21\alpha)\phi + (14+7\alpha)}{\phi^3 - 7\phi^2 + 7\phi + 7}, \tag{4.34}$$

in which we chose the degree-1 function ϕ on $\mathcal{X}/D_8$ so that $j \in \mathbb{Q}(\phi)$:

$$\begin{aligned} j &= 64 \frac{\big(\phi(\phi^2+7)(\phi^2-7\phi+14)(5\phi^2-15\phi-7)\big)^3}{(\phi^3-7\phi^2+7\phi+7)^7} \\ &= 12^3 + 56^2 \frac{(\phi-3)(2\phi^4-14\phi^3+21\phi^2+28\phi+7)P^2(\phi)}{(\phi^3-7\phi^2+7\phi+7)^7}, \end{aligned} \tag{4.35}$$

where $P(\phi)$ is the polynomial

$$P(\phi) = (\phi^4 - 14\phi^2 + 56\phi + 21)(\phi^4 - 7\phi^3 + 14\phi^2 - 7\phi + 7). \tag{4.36}$$

In the modular setting ϕ parametrizes elliptic curves E such that the Galois action on $E[7]$ is contained in a subgroup $D_8 \subset G$, i.e. by the normalizer of a nonsplit Cartan subgroup; we thus refer to the ϕ-line as the modular curve $X_n(7)$.

---

[18] Note that, since $F \supseteq k$, any $\gamma \in \mathrm{Gal}(\bar{F})/F$ must take ζ to one of ζ, ζ^2, ζ^4; thus the determinant of its action on $E[7]$ is a square in $\mathbb{F}_7^*$. Thus γ acts on $E[7]$ by a scalar multiple of a unimodular $\mathbb{F}_7$-linear transformation of $E[7]$, and may be regarded as an element of $\mathrm{PSL}_2(\mathbb{F}_7) \cong G$.

Kenku [1985] used this curve to obtain a novel proof of the Stark–Heegner theorem, which states that the only quadratic imaginary fields with unique factorization are $\mathbb{Q}(\sqrt{D})$ with $D = -3, -4, -7, -8, -11, -19, -43, -67, -163$. Let $F = \mathbb{Q}(\sqrt{D})$ be a quadratic imaginary field of discriminant $D < 0$ and class number 1. There is then an elliptic curve $E/\mathbb{Q}$ with CM by O_F, unique up to $\bar{\mathbb{Q}}$-isomorphism. Assume that the prime 7 is inert in F; this certainly happens if $|D| > 28$, else the prime(s) above 7 in F cannot be principal. (The fields with $D = -4, -8, -11$ also satisfy this condition.) Then the action of O_F on $E[7]$ gives $E[7]$ the structure of a one-dimensional vector space over $\mathbb{F}_{49}$, and $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ must respect this structure. Thus E yields a rational point of $X_n(7)$. But this point is constrained by the condition that $j_E \in \mathbb{Z}$. That is, $\phi = \phi(E)$ must be a rational number such that $j(\phi)$, given by (4.35), is an integer. Writing $\phi = m/n$ in lowest terms, we find $j(\phi) = A(m,n)/B(m,n)$ with A, B homogeneous polynomials of degree 21 without common factors. Thus $\gcd(A(m,n), B(m,n))$ is bounded given $\gcd(m,n) = 1$; one may calculate that this gcd is a factor of 56^7, and thus that $m^3 - 7m^2n + 7mn^2 + 7n^3$ divides 56. Thus if m, n are at all large then m/n must be a very good rational approximation to one of the roots $3 + 4\cos 2a\pi/7$ ($a \in \mathbb{F}_7^*$) of $\phi^3 - 7\phi^2 + 7\phi + 7$. In the present case Kenku was able to list all $\phi \in \mathbb{Q}$ such that $j(\phi) \in \mathbb{Z}$ using Nagell's list [1969] of the solutions of $x + y = 1$ in units x, y of K_+. The list can also be obtained from general bounds on rational approximation, provided all the constants are given explicitly as they are in [Bugeaud and Győry 1996]. For our specific problem of approximating elements of $K_+ \setminus \mathbb{Q}$, much better results are available, which make the computation easily tractable; for instance Michael Bennett reports that the methods of [Bennett 1997] yield the bound $\left|\cos(\pi/7) - p/q\right| > 0.099q^{-7/3}$ for all nonzero $p, q \in \mathbb{Z}$, which is more than enough to find all solutions of $|m^3 - 7mn^2 + 7mn^2 + n^3| \le 56$. We find that the list of integral points on $X_n(7)$, however obtained, consists of the points with

$$\phi \in \left\{0, \infty, 1, -1, 2, 3, 5, -\tfrac{3}{5}, 7, \tfrac{7}{3}, \tfrac{11}{2}, \tfrac{19}{9}\right\}. \tag{4.37}$$

Of the resulting integral values of $j(\phi)$, the first eight are j-invariants of CM elliptic curves, with discriminant -3, -8, -11, -16, -67, -4, -43, -163 respectively. (The discriminant -3 occurs even though 7 is split in $\mathbb{Q}(\sqrt{-3}\,)$ thanks to the cube roots of unity in $\mathbb{Q}(\sqrt{-3}\,)$, which yield extra automorphisms of a curve of j-invariant zero; $D = -16$ occurs because the order $\mathbb{Z}[2i] \in \mathbb{Q}(i)$ still has unique factorization.) It is easy to check that none of the remaining four values $j = 10^3 7^5$, $2^{15} 7^5$, $2^6 11^3 23^3 149^3 269^3$, $2^9 17^6 19^3 29^3 149^3$ can be the j-invariants of a CM curve, and this completes Kenku's proof that the list of imaginary quadratic fields of class number 1 is complete.

[We remark that Siegel [1968] had already given a similar proof of the Stark–Heegner theorem using $X_n(5)$ together with the condition that j_E is a cube, which is tantamount to using the degree-30 cover of $X(1)$ by $X_n(15)$. An amusing feature of Siegel's argument which I have not seen mentioned elsewhere is that the Diophantine equation for an integral point on $X_n(15)$ is equivalent to the condition that a Fibonacci number be a perfect cube, and thus that Siegel in effect reduced the Stark–Heegner theorem to the fact that the only such Fibonacci numbers are $0, \pm 1, \pm 8$.]

What of the four discriminants $D = -3, -12, -19, -27$ of imaginary quadratic orders with unique factorization in which 7 splits? Let E be an elliptic curve with CM by the order of discriminant $-D$. The primes above 7 yield a distinguished pair of 7-element subgroups of E, which must be respected by the Galois group. Thus j_E lifts to a rational point on the quotient of $X(7)$ by the normalizer of the *split Cartan group* of diagonal matrices. In our case the split Cartan group is $\langle h \rangle$, and its normalizer is $\langle h, s \rangle$, so we know these quotient curves already. Since S_4 contains the normalizers of both the split and the non-split Cartan groups (note that $p = 7$ is the largest case in which $\mathrm{PSL}_2(\mathbb{F}_p)$ has a proper subgroup containing Cartan normalizers of both kinds), the j-invariant of a CM curve lifts to a rational point of $X(7)/S_4$ in both the split and inert cases. These points (necessarily rational only over k, since $X(7)/S_4$ is not defined over $\mathbb{Q}$) are as follows:

D	-3	-4	-8	-11	-12
x	$2+\alpha, 3+3\alpha, -3-2\alpha$	$-4-2\alpha$	$2+3\alpha$	$5+2\alpha$	$-3+\alpha$

D	-16	-19	-27	-43	-67	-163
x	$6+4\alpha$	$5-2\alpha$	$-3+6\alpha$	$-3-14\alpha$	$42+13\alpha$	$-283-182\alpha$

This accounts for all but two of the thirteen rational j-invariants. The remaining rational j's have $D = -7$ and $D = -28$; these are the j-invariants -15^3, 255^3 of the curves E_k, E_k', for which 7 is ramified in the CM field $\mathbb{Q}(\sqrt{D})$, a.k.a. k. These two j's lift to rational points not on $X(7)/S_4$ but on $X_0(7)$, in fact to the fixed points $j_7 = -7$ and $j_7 = +7$ of the involution w_7.

4.4. $\mathcal{X}$ as a Shimura curve. Our identification of $\mathcal{X}$ with $X_0(7) = \mathcal{H}^*/\Gamma_0(7)$ identifies $\Gamma_0(7)$ with the fundamental group not of $\mathcal{X}$ but of $\mathcal{X}$ punctured at the 24-point orbit. We have seen already that in the hyperbolic uniformization of $\mathcal{X}$ the fundamental group $\pi_1(\mathcal{X})$ becomes a normal subgroup of the triangle group $G_{2,3,7}$. Remarkably this too is an arithmetic group: let

$$c = \zeta + \zeta^{-1} = 2\cos(2\pi/7), \tag{4.38}$$

so $O_{K_+} = \mathbb{Z}[c]$; then there exist matrices $i, j \in \mathrm{GL}_2(\mathbb{R})$ such as $c^{1/2}\left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)$ and $c^{1/2}\left(\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right)$ with

$$i^2 = j^2 = c \cdot \mathbf{1}, \quad ij = -ji \tag{4.39}$$

(this determines i, j uniquely up to $\mathrm{GL}_2(\mathbb{R})$ conjugation) such that $G_{2,3,7}$ consists of the images in $\mathrm{PSL}_2(\mathbb{R})$ of $\mathbb{Z}[c]$-linear combinations of $\mathbf{1}, i, j', ij'$ whose determinant equals 1. Here

$$j' := \tfrac{1}{2}\bigl(1 + ci + (c^2 + c + 1)j\bigr), \tag{4.40}$$

and the determinant of $a_1\mathbf{1} + a_2 i + a_3 j + a_4 ij$ $(a, b, c, d \in \mathbb{R})$ is

$$a_1^2 - ca_2^2 - ca_3^2 + c^2 a_4^2. \tag{4.41}$$

For instance, $G_{2,3,7}$ is generated by the images in $\mathrm{PSL}_2(\mathbb{R})$ of

$$\begin{gathered} g_2 := ij/c, \qquad g_3 := \tfrac{1}{2}\bigl(1 + (c^2 - 2)j + (3 - c^2)ij\bigr), \\ g_7 := \tfrac{1}{2}\bigl(c^2 + c - 1 + (2 - c^2)i + (c^2 + c - 2)ij\bigr), \end{gathered} \tag{4.42}$$

with $g_2^2 = g_3^3 = g_7^7 = -1$ and $g_2 = g_7 g_3$. (Note that "$g_2 = ij/c$" is legitimate since c is a unit.] Shimura [1967] found that the quotients of $\mathcal{H}$ by arithmetic groups or their congruence subgroups also have modular interpretations, analogous to the interpretation of $\mathcal{H}^*/\Gamma(N)$ as the moduli space for elliptic curves with full level-N structure. The objects parametrized by Shimura's modular curves are more complicated than elliptic curves; for instance $\mathcal{X}$ and $\mathcal{X}/G$ parametrize families of principally polarized abelian varieties of dimension 6. These abelian sixfolds can be described precisely, but there is as yet no hope of presenting them explicitly enough to derive formulas for the sixfold parametrized by a given point of $\mathcal{X}/G$ or of $\mathcal{X}$. Still these curves hold a place in number theory comparable to that of the classical modular curves coming from congruence subgroups of $\Gamma(1)$, and limited computational investigation of these curves is now feasible (see for instance [Elkies 1998b]). For the our present purposes we content ourselves with describing the specific arithmetic groups and moduli problems connected with the Klein quartic, referring the reader to [Vignéras 1980] for the arithmetic of quaternion algebras over number fields in general, and to [Vignéras 1980; Shimura 1967] for their associated Shimura modular curves.

The K_+-algebra A generated by i, j is a *quaternion algebra* over K_+: a simple associative algebra with unit, containing K_+, such that K_+ is the center of A and $\dim_{K_+} \mathsf{A} = 4$. The ring $\mathcal{O} = O_{K_+}[i, j'] \subset \mathsf{A}$ is a maximal order in A. For each of the three real places v of K_+ we may form a quaternion algebra over $\mathbb{R}$ by tensoring A with $(K_+)_v \cong \mathbb{R}$. It is known that a quaternion algebra over $\mathbb{R}$ is isomorphic with either the algebra $M_2(\mathbb{R})$ of 2×2 real matrices, or with the Hamilton quaternions $\mathbb{H}$. We have seen that in our chosen real embedding of K_+, taking c to $2\cos(2\pi/7)$, the algebra $\mathsf{A} \otimes_{K_+} (K_+)_v$ is $M_2(\mathbb{R})$; for the other two places, in which c is $2\cos(4\pi/7)$ and $2\cos(8\pi/7)$, that algebra is isomorphic with $\mathbb{H}$ because then $i^2, j^2 < 0$. It is known that if a quaternion algebra over a number field F becomes isomorphic with $M_2(\mathbb{R})$ over at least one of F's real places then the maximal order $\mathcal{O}$ is unique up to conjugation in the algebra; moreover, that if (as in our case) there is exactly one such place and F is totally real then the group of units of norm 1 in $\mathcal{O}^* \hookrightarrow \mathrm{GL}_2(\mathbb{R})$ yields a co-compact subgroup $\Gamma \cong \mathcal{O}^*/\{\pm 1\}$ of $\mathrm{PSL}_2(\mathbb{R})$, and thus a compact Riemann surface "$X(1)$" $:= \mathcal{H}/\Gamma$, except in the classical case of the algebra $M_2(\mathbb{Q})$ over $\mathbb{Q}$. Since all maximal orders are conjugate, the resulting curve does not depend on the choice of maximal order $\mathcal{O}$. As a modular curve, "$X(1)$" parametrizes principally polarized abelian varieties of dimension $2[K:\mathbb{Q}]$ ($=6$ in our case) with endomorphisms by $\mathcal{O}$. This means that the curve "$X(1)$", though constructed transcendentally, is defined over some number field; in our case that field may even be taken to be $\mathbb{Q}$ thanks to the facts that K_+ has unique factorization and is Galois over $\mathbb{Q}$. Since for us $\Gamma \cong G_{2,3,7}$, this curve is rational: the quotient of $\mathcal{H}$ by any triangle group has genus zero.

Our quaternion algebra A over K_+ has the remarkable property that, for each *finite* place v of K_+, the quaternion algebra $\mathsf{A} \otimes_{K_+} (K_+)_v$ over $(K_+)_v$ is isomorphic with $M_2((K_+)_v)$. (In other words, A is *unramified* at each finite prime v.) Using this isomorphism, one may define arithmetic subgroups of Γ and modular curves covering "$X(1)$" analogous to the classical modular curves $X(N)$, $X_0(N)$ etc. For instance if $\wp$ is a prime of O_K then the units of $\mathcal{O}$ congruent to 1 mod $\wp$ constitute a normal subgroup of $\mathcal{O}^*$ that maps to a normal subgroup $\Gamma(\wp)$ of Γ. Thanks to the isomorphism of $\mathsf{A} \otimes_{K_+} (K_+)_v$ with $M_2((K_+)_v)$ we have $\Gamma/\Gamma(\wp) \cong \mathrm{PSL}_2(k_\wp)$ [where $k_\wp$ is the residue field $O_{K_+}/\wp$ of $\wp$]. The Riemann surface "$X(\wp)$" $:= \mathcal{H}/\Gamma(\wp)$ is then a normal cover of "$X(1)$" with Galois group $\mathrm{PSL}_2(k_\wp)$. This too is a Shimura modular curve, parametrizing principally polarized abelian sixfolds with endomorphisms by $\mathcal{O}$ and complete level-$\wp$ structure — this last makes sense because $O_{K_+} \subset \mathcal{O}$ acts on the sixfold so we may speak about the sixfold's $\wp$-torsion points. The isomorphism $\Gamma/\Gamma(\wp) \cong \mathrm{PSL}_2(k_\wp)$ lets us define groups $\Gamma_0(\wp), \Gamma_1(\wp)$ intermediate between Γ and $\Gamma(\wp)$, and thus Shimura modular curves "$X_0(\wp)$" and "$X_1(\wp)$", which parametrize $\mathcal{O}$-sixfolds with partial level-$\wp$ structure. The curves "$X(\wp)$", "$X_0(\wp)$" and "$X_1(\wp)$" are defined over K_+, and even over $\mathbb{Q}$ if $\wp$ is Galois-stable. Note that the Galois-stable primes of K_+ are those that lie over an inert rational prime, i.e. a prime $\equiv \pm 2$ or ± 3 mod 7, and the prime $\wp_7 = (2-c)$ lying over the ramified prime 7.

We remarked already that Hurwitz curves come from normal subgroups of $G_{2,3,7}$. Shimura observed [1967, p. 83] that since each of the groups $\Gamma(\wp)$ is a normal subgroup of Γ, and $\Gamma \equiv g_{2,3,7}$, the resulting curves "$X(\wp)$" are Hurwitz curves. In particular "$X(\wp_7)$" is a Hurwitz curve of genus 3. We already know what this means: "$X(\wp_7)$" is none other than the Klein quartic $\mathcal{X}$. Furthermore, its fundamental group $\pi_1(\mathcal{X})$ is the congruence subgroup of Γ consisting of the images in $\mathrm{PSL}_2(\mathbb{R})$ of $\mathbb{Z}[c]$-linear combinations $a_1\mathbf{1} + a_2 i + a_3 j' + a_4 ij'$ of norm 1 with $2-c$ dividing a_2, a_3, a_4.

[The four Hurwitz curves of the next smallest genera also arise as "$X(\wp)$" for primes $\wp$ of K_+: the prime above 2 yields the Fricke–Macbeath curve [Fricke 1899; Macbeath 1965] of genus 7 and automorphism group (P)$\mathrm{SL}_2(\mathbb{F}_8)$, and the primes above 13 yield three curves of genus 14 with automorphisms by $\mathrm{PSL}_2(\mathbb{F}_{13})$ first found by Shimura. The next two Hurwitz curves have genus 17 and come from non-arithmetic quotients of $G_{2,3,7}$. See [Conder 1990] for more information on the groups that can arise as automorphism groups of Hurwitz curves, and [Conder 1987] for the list of all such groups of order less than 10^6.]

The quotient curves $X_0(7)$, $X_1(7)$, $X_0(49)$ of $\mathcal{X}$ now reappear as Shimura modular curves "$X_0(\wp_7)$", "$X_1(\wp_7)$", "$X_0(\wp_7^2)$". These curves have involutions $w_{\wp_7}$ and $w_{\wp_7^2}$ analogous to the Fricke involutions of the classical modular curves. However, the involutions of "$X_0(\wp_7)$" and "$X_1(\wp_7)$" are not the same as the involutions of the same quotients of $\mathcal{X}$ when considered as the classical modular curves $X_0(7)$ and $X_1(7)$. For instance, on $X_0(7)$ the involution $w_7 : j_7 \leftrightarrow 49/j_7$ switched the two cusps $j_7 = 0, \infty$, and also the elliptic points of order 3, at which

$$j_7^2 + 13j_7 + 49 = 0.$$

On "$X_0(\wp_7)$", the elliptic points of order 3 remain the same and are still switched by $w_{\wp_7}$; but there are no cusps — instead, the simple pole $j_7 = \infty$ of j is the unique elliptic point of order 7 of "$X_0(\wp_7)$", and must thus be fixed by $w_{\wp_7}$. Therefore $w_{\wp_7}$ takes j_7 not to $49/j_7$ but to $-13 - j_7$. In this setting the three Fricke involutions of "$X_1(\wp_7)$" are defined over $\mathbb{Q}$, and take d to $1-d$, $1/d$, and $d/(d-1)$.

We have seen already that the Fermat curve $\mathcal{F}_7$ is an unramified cover of $\mathcal{X}$. It follows that $\pi_1(\mathcal{F}_7)$ is a subgroup of $\pi_1(\mathcal{X})$, and thus of $G_{2,3,7}$. That subgroup obligingly turns out to be a congruence subgroup, with the result that $\mathcal{F}_7$, like $\mathcal{X}$, is a Shimura modular curve. That subgroup — call it Γ_7 — is intermediate between $\Gamma(\wp_7)$ and $\Gamma(\wp_7^2)$, and may be described as follows: under an identification of $\Gamma/\Gamma(\wp_7^2)$ with $\mathrm{PSL}_2(O_{K_+}/\wp_7^2)$, the group $\Gamma_7/\Gamma(\wp_7^2)$ consists of matrices congruent to the identity mod $\wp$ whose bottom left entry vanishes. Clearly Γ_7, thus defined, contains $\Gamma(\wp_7)$ as a normal subgroup of index 7, so $\mathcal{H}/\Gamma_7$ is a degree-7 unramified cyclic cover of $\mathcal{X}$. This is not yet enough to identify $\mathcal{H}/\Gamma_7$ with $\mathcal{F}_7$, but we obtain more automorphisms of $\mathcal{H}/\Gamma_7$ by observing that $\Gamma_0(\wp_7)$ is also a normal subgroup. Thus the quotient group $\Gamma_0(\wp_7)/\Gamma_7$ acts on $\mathcal{H}/\Gamma_7$. This group of automorphisms contains as an index-3 normal subgroup $\Gamma_1(\wp_7)/\Gamma_7$, which is an elementary abelian group of order 7^2. The quotient of $\mathcal{H}/\Gamma_7$ by this subgroup is the genus-zero curve $\mathcal{H}/\Gamma_1(\wp_7) =$ "$X_1(\wp_7)$", which we have already described as $X_1(7) = \mathcal{X}/\langle h\rangle$; and the ramification behavior of this quotient map $(\mathcal{H}/\Gamma_7)\to$"$X_1(\wp_7)$" *does* suffice to identify $\mathcal{H}/\Gamma_7$ with $\mathcal{F}_7$. The 147-element group $\Gamma_0(\wp_7)/\Gamma_7$ is then an index-2 subgroup of $\mathrm{Aut}(\mathcal{F}_7)$, generated by diagonal 3×3 matrices and cyclic coordinate permutations; extending $\Gamma_0(\wp_7)$ by $w_{\wp_7}$ yields the full group of automorphisms of $\mathcal{F}_7$.

References

[Adler 1981] A. Adler, "Some integral representations of $\mathrm{PSL}_2(\mathbf{F}_p)$ and their applications", *J. Algebra* **72**:1 (1981), 115–145.

[Adler 1997] A. Adler, "The Mathieu group M_{11} and the modular curve $X(11)$", *Proc. London Math. Soc.* (3) **74**:1 (1997), 1–28.

[Arbarello et al. 1985] E. Arbarello, M. Cornalba, P. A. Griffiths, and J. Harris, *Geometry of algebraic curves, I*, Grundlehren der mathematischen Wissenschaften **267**, Springer, New York, 1985.

[Bennett 1997] M. A. Bennett, "Effective measures of irrationality for certain algebraic numbers", *J. Austral. Math. Soc. Ser. A* **62**:3 (1997), 329–344.

[Benson 1993] D. J. Benson, *Polynomial invariants of finite groups*, London Math. Soc. Lecture Note Series, Cambridge University Press, Cambridge, 1993.

[Birch and Kuyk 1975] B. J. Birch and W. Kuyk (editors), *Modular functions of one variable, IV*, edited by B. J. Birch and W. Kuyk, Lecture Notes in Math. **476**, Springer, Berlin, 1975.

[Bourbaki 1968] N. Bourbaki, *Groupes et algèbres de Lie, IV–VI*, Actualités scientifiques et industrielles **1337**, Hermann, Paris, 1968. Reprinted by Masson, Paris, 1981.

[Bugeaud and Győry 1996] Y. Bugeaud and K. Győry, "Bounds for the solutions of unit equations", *Acta Arith.* **74**:1 (1996), 67–80.

[Buser and Sarnak 1994] P. Buser and P. Sarnak, "On the period matrix of a Riemann surface of large genus", *Invent. Math.* **117**:1 (1994), 27–56. With an appendix by J. H. Conway and N. J. A. Sloane.

[Buser et al. 1994] P. Buser, J. Conway, P. Doyle, and K.-D. Semmler, "Some planar isospectral domains", *Internat. Math. Res. Notices* **1994**:9 (1994), 391–399.

[Clemens 1980] C. H. Clemens, *A scrapbook of complex curve theory*, Plenum, New York, 1980.

[Conder 1987] M. Conder, "The genus of compact Riemann surfaces with maximal automorphism group", *J. Algebra* **108**:1 (1987), 204–247.

[Conder 1990] M. Conder, "Hurwitz groups: a brief survey", *Bull. Amer. Math. Soc. (N.S.)* **23**:2 (1990), 359–370.

[Conway et al. 1985] J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, and R. A. Wilson, *Atlas of finite groups*, Oxford University Press, Oxford, 1985.

[Coolidge 1931] J. L. Coolidge, *A treatise on algebraic plane curves*, Clarendon, Oxford, 1931.

[Cremona 1992] J. E. Cremona, *Algorithms for modular elliptic curves*, Cambridge University Press, Cambridge, 1992.

[Darmon and Granville 1995] H. Darmon and A. Granville, "On the equations $z^m = F(x,y)$ and $Ax^p + By^q = Cz^r$", *Bull. London Math. Soc.* **27**:6 (1995), 513–543.

[Dickson 1934] L. E. Dickson, *History of the theory of numbers, II: Diophantine analysis*, Stechert and Co., New York, 1934.

[Ekedahl and Serre 1993] T. Ekedahl and J.-P. Serre, "Exemples de courbes algébriques à jacobienne complètement décomposable", *C. R. Acad. Sci. Paris Sér. I Math.* **317**:5 (1993), 509–513.

[Elkies 1994] N. D. Elkies, "Mordell-Weil lattices in characteristic 2, I: Construction and first properties", *Internat. Math. Res. Notices* **1994**:8 (1994), 343–361.

[Elkies 1998a] N. D. Elkies, "Elliptic and modular curves over finite fields and related computational issues", pp. 21–76 in *Computational perspectives on number theory*

(Chicago, 1995), edited by D. A. Buell and J. T. Teitelbaum, AMS/IP Stud. Adv. Math., Amer. Math. Soc., Providence, RI, 1998.

[Elkies 1998b] N. D. Elkies, "Shimura curves computations", pp. 1–47 in *Algorithmic number theory* (ANTS-III: Portland, 1998), edited by J. Buhler, Lecture Notes in Computer Science **1423**, Springer, New York, 1998.

[Faltings 1983] G. Faltings, "Endlichkeitssätze für abelsche Varietäten über Zahlkörpern", *Invent. Math.* **73**:3 (1983), 349–366. Erratum in **75** (1984), 381.

[Faltings 1991] G. Faltings, "Diophantine approximation on abelian varieties", *Ann. of Math.* (2) **133**:3 (1991), 549–576.

[Fricke 1899] R. Fricke, "Über eine einfache Gruppe von 504 Operationen", *Math. Annalen* **52** (1899), 321–339.

[Fulton and Harris 1991] W. Fulton and J. Harris, *Representation theory*, Graduate Texts in Math. **129**, Springer, New York, 1991.

[Genocchi 1864] A. Genocchi, "Intorno all'equazione $x^7 + y^7 + z^7 = 0$", *Annali di Mat. Pura ed Applicata* **6** (1864), 287–288.

[Gordon et al. 1992] C. Gordon, D. L. Webb, and S. Wolpert, "One cannot hear the shape of a drum", *Bull. Amer. Math. Soc. (N.S.)* **27**:1 (1992), 134–138.

[Gross 1978] B. H. Gross, "On the periods of abelian integrals and a formula of Chowla and Selberg", *Invent. Math.* **45**:2 (1978), 193–211. With an appendix by David E. Rohrlich.

[Gross 1990] B. H. Gross, "Group representations and lattices", *J. Amer. Math. Soc.* **3**:4 (1990), 929–960.

[Gross and Rohrlich 1978] B. H. Gross and D. E. Rohrlich, "Some results on the Mordell-Weil group of the Jacobian of the Fermat curve", *Invent. Math.* **44**:3 (1978), 201–224.

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Math. **52**, Springer, New York, 1977.

[Hirzebruch 1983] F. Hirzebruch, "Arrangements of lines and algebraic surfaces", pp. 113–140 in *Arithmetic and geometry, II*, edited by M. Artin and J. Tate, Progr. Math. **36**, Birkhäuser, Boston, 1983. Reprinted as #69 (pp. 679–706) of his *Gesammelte Abhandlungen* vol. 2, Springer, 1987.

[Hoffmann 1991] D. W. Hoffmann, "On positive definite Hermitian forms", *Manuscripta Math.* **71**:4 (1991), 399–429.

[Hurwitz 1893] A. Hurwitz, "Über algebraische Gebilde mit eindeutigen Transformationen in sich", *Math. Annalen* **41** (1893), 403–442.

[Kemper 1996] G. Kemper, "A constructive approach to Noether's problem", *Manuscripta Math.* **90**:3 (1996), 343–363.

[Kenku 1985] M. A. Kenku, "A note on the integral points of a modular curve of level 7", *Mathematika* **32**:1 (1985), 45–48.

[Klein 1879a] F. Klein, "Ueber die Erniedrigung der Modulargleichungen", *Math. Annalen* **14** (1879), 417–427. Reprinted as [Klein 1923, LXXXIII, pp. 76–89].

[Klein 1879b] F. Klein, "Ueber die Transformationen siebenter Ordnung der elliptischen Funktionen", *Math. Annalen* **14** (1879), 428–471. Reprinted as [Klein 1923, LXXXIV, pp. 90–136]. Translated in this collection.

[Klein 1923] F. Klein, *Gesammelte Mathematische Abhandlungen, 3: Elliptische Funktionen* etc., edited by R. Fricke et al., Springer, Berlin, 1923. Reprinted by Springer, 1973.

[Kneser 1967] M. Kneser, "Über die Ausnahme-Isomorphismen zwischen endlichen klassischen Gruppen", *Abh. Math. Sem. Univ. Hamburg* **31** (1967), 136–140.

[Kubert and Lang 1981] D. S. Kubert and S. Lang, *Modular units*, Grundlehren der mathematischen Wissenschaften **244**, Springer, New York, 1981.

[Macbeath 1965] A. M. Macbeath, "On a curve of genus 7", *Proc. London Math. Soc.* (3) **15** (1965), 527–542.

[MacWilliams and Sloane 1977] F. J. MacWilliams and N. J. A. Sloane, *The theory of error-correcting codes*, North-Holland Mathematical Library **16**, North-Holland, Amsterdam, 1977.

[Mazur 1986] B. Mazur, "Arithmetic on curves", *Bull. Amer. Math. Soc.* (*N.S.*) **14**:2 (1986), 207–259.

[Nagell 1960] T. Nagell, "The Diophantine equation $x^2 + 7 = 2^n$", *Ark. Mat.* **4** (1960), 185–187.

[Nagell 1969] T. Nagell, "Sur un type particulier d'unités algébriques", *Ark. Mat.* **8** (1969), 163–184.

[Perlis 1977] R. Perlis, "On the equation $\zeta_K(s) = \zeta_{K'}(s)$", *J. Number Theory* **9**:3 (1977), 342–360.

[Poonen 1996] B. Poonen, "Computational aspects of curves of genus at least 2", pp. 283–306 in *Algorithmic number theory: Second International Symposium* (Talence, 1996), edited by H. Cohen, Lecture Notes in Comput. Sci. **1122**, Springer, Berlin, 1996.

[Selberg and Chowla 1967] A. Selberg and S. Chowla, "On Epstein's zeta-function", *J. Reine Angew. Math.* **227** (1967), 86–110.

[Serre 1967] J.-P. Serre, "Complex multiplication", pp. 292–296 in *Algebraic Number Theory* (Brighton, 1965), edited by J. W. S. Cassels and A. Fröhlich, Thompson and Academic Press, Washington, D.C., and London, 1967. Reprinted as #76, (pp. 455–459) of his *Oeuvres*, vol. 2, Springer, Berlin, 1986.

[Serre 1973] J.-P. Serre, *A course in arithmetic*, Graduate Texts in Math. **7**, Springer, New York, 1973. Translation of *Cours d'arithmétique*, Presses univ. de France, Paris, 1970.

[Serre 1983a] J.-P. Serre, "Nombres de points des courbes algébriques sur $\mathbf{F}_q$", pp. Exp. No. 22, 8 in *Séminaire de théorie des nombres de Bordeaux* (Talence, 1982/1983), Univ. Bordeaux I, Talence, 1983. Reprinted as #129 (pp. 664–668) of his *Oeuvres*, vol. 3, Springer, Berlin, 1986.

[Serre 1983b] J.-P. Serre, "Sur le nombre des points rationnels d'une courbe algébrique sur un corps fini", *C. R. Acad. Sci. Paris Sér. I Math.* **296**:9 (1983), 397–402. Reprinted as #128 (pp. 658–663) of his *Oeuvres*, vol. 3, Springer, Berlin, 1986.

[Serre 1984] J.-P. Serre, "Résumés des cours de 1983–1984", pp. 79–83 in *Annuaire*, Collège de France, Paris, 1984. Reprinted as #132 (pp. 701–705) of his *Oeuvres*, vol. 3, Springer, Berlin, 1986.

[Shephard and Todd 1954] G. C. Shephard and J. A. Todd, "Finite unitary reflection groups", *Canadian J. Math.* **6** (1954), 274–304.

[Shimura 1967] G. Shimura, "Construction of class fields and zeta functions of algebraic curves", *Ann. of Math.* (2) **85** (1967), 58–159.

[Siegel 1968] C. L. Siegel, "Zum Beweis des Starkschen Satzes", *Invent. Math.* **5** (1968), 180–191.

[Silverman 1986] J. H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Math. **106**, Springer, New York, 1986.

[Stark 1973] H. M. Stark, "On the Riemann hypothesis in hyperelliptic function fields", pp. 285–302 in *Analytic number theory* (St. Louis, MO, 1972), edited by H. G. Diamond, Proc. Sympos. Pure Math. **24**, Amer. Math. Soc., Providence, 1973.

[Stichtenoth 1973] H. Stichtenoth, "Über die Automorphismengruppe eines algebraischen Funktionenkörpers von Primzahlcharakteristik", *Arch. Math.* (*Basel*) **24** (1973), 527–544, 615–631.

[Sunada 1985] T. Sunada, "Riemannian coverings and isospectral manifolds", *Ann. of Math.* (2) **121**:1 (1985), 169–186.

[Suzuki 1982] M. Suzuki, *Group theory, I*, Grundlehren der mathematischen Wissenschaften, Springer, Berlin, 1982. Translated from the Japanese by the author.

[Tate 1974] J. T. Tate, "The arithmetic of elliptic curves", *Invent. Math.* **23** (1974), 179–206.

[Thompson 1976] J. G. Thompson, "Finite groups and even lattices", *J. Algebra* **38**:2 (1976), 523–524.

[Vélu 1971] J. Vélu, "Isogénies entre courbes elliptiques", *C. R. Acad. Sci. Paris Sér. A-B* **273** (1971), A238–A241.

[Vignéras 1980] M.-F. Vignéras, *Arithmétique des algèbres de quaternions*, Lecture Notes in Math. **800**, Springer, Berlin, 1980.

[Weil 1964] A. Weil, "Sur certains groupes d'opérateurs unitaires", *Acta Math.* **111** (1964), 143–211.

Noam D. Elkies
Department of Mathematics
Harvard University
Cambridge, MA 02138
United States
elkies@math.harvard.edu

The Eightfold Way
MSRI Publications
Volume **35**, 1998

Hurwitz Groups and Surfaces

A. MURRAY MACBEATH

Abstract. Hurwitz not only gave an upper bound for the number of automorphisms of a compact Riemann surface of genus greater than 2, but also gave a characterization of which finite groups could be groups of automorphisms achieving this bound. In practice, however, the identification of such groups and of the surfaces they act on is difficult except in special cases. We survey what is known.

1. How I Got Started on Hurwitz Groups

One day in the late 1950's, rereading Siegel's article [1945] entitled "Some remarks on discontinuous groups", I was struck by his proof that the smallest area of fundamental region for a Fuchsian group is $\pi/21$.

Siegel notes the remarkable similarity between the arithmetic used in his proof and the arithmetic in Hurwitz's proof that a curve of genus $g \geq 2$ has no more than $84(g-1)$ birational self-transformations. That, he said, is not surprising because of the theory of uniformization. That was all — no indication where to find Hurwitz's paper, at that time unknown to me. (Siegel is one of my heroes, but, it must be confessed, he was not very good at citing references.)

I did know about uniformization, and I made that connection at once. However, I had some trouble tracking down Hurwitz's theorem. Finally, thanks to the late Professor W. L. Edge, I read Hurwitz's paper [1893], which invoked Klein's surface as an example to show that his bound was attained. So at last, by a very tortuous path, I unearthed this chapter of mathematics, which has fascinated me ever since.

Hurwitz left open the question whether there was any other surface with the maximum number $84(g-1)$ of automorphisms, as we now call them. Only one other such surface was found, by Fricke, in the sixty years to 1961. My own first contribution [Macbeath 1961] was a proof that there are infinitely many of them.

My research changed direction when I became aware of Klein's curve and Hurwitz's theorem. I was driven to think more and more about Riemann surfaces with many automorphisms. It was natural to progress to Riemann surfaces in general and to Teichmüller spaces. Friends and colleagues, whether their first interest might be geometry, algebra, analysis or number theory, found points of contact with this work. It is a truly central piece of mathematics.

The Klein surface is the Riemann surface of the algebraic curve with equation, in homogeneous coordinates $x : y : z$,

$$x^3y + y^3z + z^3x = 0. \tag{1}$$

Klein [1879] showed that it is mapped on itself by 168 analytic transformations. Since the equation is real, the surface is also mapped on itself by complex conjuation, which can be composed with the analytic maps to give a further 168 antianalytic mappings, yielding a group of order 336. Klein concentrated his attention on the subgroup of index 2 and order 168.

2. Klein

That group is the second smallest simple noncommutative group. (From now on we will write "simple group" for "simple noncommutative group".) It belongs to two infinite families, $\mathrm{PSL}(2,7) \cong \mathrm{PSL}(3,2)$. For Klein it would certainly have been $\mathrm{PSL}(2,7)$ (if the notation had been invented), because he approached the situation — group *and* Riemann surface — by studying the modular group $\mathbf{\Gamma}(1)$ of all functions

$$z \mapsto \frac{pz+q}{rz+s}, \tag{2}$$

where $p, q, r, s \in \mathbb{Z}$, and $ps - qr = 1$. These are permutations of the upper half-plane $\mathbb{U} := \{z \in \mathbb{C} \mid i(\bar{z} - z) > 0\}$.

Since the integers are a discrete subset of the reals, $\mathbf{\Gamma}(1)$ is, in any reasonable sense, a discontinuous group of mappings. The upper half-plane is a Riemann surface, so its quotient surface $\mathbb{U}/\mathbf{\Gamma}(1)$ is also a Riemann surface — a sphere with one missing point, or *puncture*. This is a slight disappointment if we are looking for interesting Riemann surfaces. Subgroups of $\mathbf{\Gamma}(1)$ might do better.

The *congruence subgroups* $\mathbf{\Gamma}(n)$, which consist of mappings (2) such that

$$\begin{pmatrix} p & q \\ r & s \end{pmatrix} \equiv \pm \mathrm{Id} \pmod{n},$$

are the first to jump up and hit us. Being the kernel of a homomorphism, $\mathbf{\Gamma}(n)$ is a normal subgroup of $\mathbf{\Gamma}(1)$, and the factor group acts on the quotient surface as a group of automorphisms.

The quotient surfaces for $\mathbf{\Gamma}(2)$, $\mathbf{\Gamma}(3)$, $\mathbf{\Gamma}(4)$ and $\mathbf{\Gamma}(5)$ are spheres with 3, 4, 6 and 12 punctures. The factor groups include the symmetry groups of the platonic solids (tetrahedron, octahedron and icosahedron). The quotient surface of $\mathbf{\Gamma}(6)$, a torus with twelve punctures, is slightly more interesting, but the factor group $\mathbf{\Gamma}(1)/\mathbf{\Gamma}(6)$ is rather dull. Klein had studied all these groups in detail from the viewpoints of complex analysis and projective geometry.

At $\mathbf{\Gamma}(7)$, he found buried treasure. This surface has genus 3 with 24 punctures. The punctures are "removable singularities" and pose no problem, so he had a Riemann surface of genus 3 with 168 automorphisms. The quotient group is what we get when we replace the integers in (2) by their residue classes modulo 7, a group now denoted by $\mathrm{PSL}(2,7)$.

Now when we know that a surface S exists, then by general theory there is a pair of meromorphic functions x, y on S that can distinguish any one point of S from any other. The functions satisfy a polynomial identity $F(x,y)=0$, defining an algebraic curve. The curve has a vector space of abelian differentials, it has Weierstrass points and all the other good things that an algebraic curve possesses. Having found S, some of us might be content to rest on our laurels in the mere knowledge that these things exist, but Klein was made of the right stuff. He had to know what they were.

Not only did he find the equation (1) — no mean achievement from such meagre data — he also found explicitly all the biholomorphic mappings as 3×3 matrices. These define projective transformations mapping the curve on itself. By doing this he closed one of two gaps in Jordan's list [1878] of finite ternary linear groups, as Fricke points out [1926, footnote on p. 182].

It sems that Klein *began* with the differentials, and then worked out the linear mappings induced on them by the 168 automorphisms. He studied the invariants of this linear group, finding three basic invariants that are connected by equation (1).

The Riemann surface of the curve (1) is a 168-sheeted covering of the sphere, branched over three points of the sphere.

Above one of these points the 168 sheets join together in sevens to give 24 points of the surface. These are the *points of inflection.* They are also the Weierstrass points.

Above another branch point, there are 84 points of the surface, where the sheets join in twos. These are the *sextactic* points, through which pass a conic section that has six-fold contact with the curve.

Above the third branch point the sheets join in threes to give 56 points of the surface. These 56 points are the points of contact of (1) with the 28 *bitangents*, or lines that are tangent to the curve at two points.

All these facts were discovered by Klein.

The numbers 2, 3, 7 reflect the fact that the universal cover of the whole picture is the triangle group $(2,3,7)$ acting on $\mathbb{U}$. The modular group $\mathbf{\Gamma}(1)$ is the triangle group $(2,3,\infty)$. Replacing ∞ by 7 amounts to removing the removable singularities.

For more detail see [Fricke 1926, p. 182–235] or the translation of Klein's article in this volume.

3. Hurwitz

Hurwitz's paper [1893] is a bold piece of work. His aim, as the title indicates, was to study the general situation of an algebraic curve and a group $\mathbf{\Gamma}$ of automorphisms. It had been proved by Schwarz that the automorphism group of a curve of genus $g \geq 2$ is finite, so he assumed that $g \geq 2$.

His approach was topological, considering the Riemann surface of the curve as a branched covering of the quotient surface of $\mathbf{\Gamma}$-orbits. He worked out the relation between the genus of the surface, the genus of the quotient surface and the branching numbers. We now call this the *Riemann–Hurwitz relation*. From this he worked out the upper bound $84(g-1)$ mentioned by Siegel.

Among other results, Hurwitz also proved that the action of the automorphism group on the abelian differentials is faithful, and that the order of any single automorphism cannot exceed $10(g-1)$. He also showed that a finite group can be realized as a group of $84(g-1)$ automorphisms of a surface of genus $g \geq 2$ *if and only if it is generated by two elements t, u such that*

$$t^2 = u^3 = (tu)^7 = 1.$$

Such a group is now called a *Hurwitz group*. Even if we did not know anything about Fuchsian groups, we would nowadays feel forced to invent the abstract *triangle group*

$$\langle t, u \mid t^2 = u^3 = (tu)^7 = 1 \rangle \tag{3}$$

and to rephrase the result:

The Hurwitz groups are precisely the finite homomorphic images of (3).

The problem of finding surfaces with $84(g-1)$ automorphisms is now reduced to a purely group-theoretic question. Without making the connection to Riemann surfaces, G. A. Miller [1902] proved that there are infinitely many Hurwitz groups.

4. Poincaré

The introduction of Fuchsian groups by Poincaré [1882] had a strong influence on our way of thinking about Riemann surfaces. Though his work was a decade before Hurwitz's, it is quite clear that Hurwitz was writing without reference to it, and perhaps he did not know of it. Some very effective work on automorphisms, for example [Accola 1968], has been done quite recently without any mention of Fuchsian groups, using covering space theory.

For me, though, the intuitive picture gained from Fuchsian groups is all-important. I see the automorphism group as a tiling of the surface, the quotient surface being what we get when we identify matched edges of any one tile. By rolling the whole surface out on to the simply connected universal covering surface $\mathbb{U}$ we get a coarser tiling (of $\mathbb{U}$) whose matching gives the target surface of the automorphism group. Each of the coarse tiles is a mosaic of finer tiles, whose edge-matching gives the quotient surface.

All the fine tiles have the same hyperbolic area, say a, and all the coarse tiles have the same area A. The order n of the automorphism group is the number of fine tiles that fit into one coarse tile. Therefore

$$A = na.$$

This is the *Riemann–Hurwitz relation*, which, in this form, seems blindingly obvious. To use the relation effectively, we must rewrite the areas A, a in terms of algebraic invariants of the Fuchsian groups. Still, the use of Fuchsian groups makes everything more transparent.

Poincaré recognised $\mathbb{U}$ as the hyperbolic plane, which does not admit arbitrarily fine congruent tilings. Siegel's theorem is an exact quantitative expression of this. The upper half-plane $\mathbb{U}$ had been known as the target space for the modular group before Poincaré was even born, and hyperbolic (or, as it was then called, noneuclidean) geometry had been studied for its own sake. With the metric

$$\frac{|dz|}{y},$$

$\mathbb{U}$ had been used as a model for hyperbolic geometry by Beltrami (see, e.g., [Stillwell 1996]). Regarded by many mathematicians as a gigantic counterexample designed to show that Euclid's geometry could not be deduced without the parallel postulate, hyperbolic geometry had to wait until Poincaré to be synthesized with the modular figure and admitted to mathematical respectability.

Klein's approach to his curve, involving the modular group, is much closer to Poincaré than to Hurwitz. Even though the groundwork was done without specific mention of general Fuchsian groups, the ideas were in the air, just waiting for someone like Poincaré to crystallize them.

For historical and mathematical insight on the emergence of hyperbolic geometry from the shadows see the collection [Stillwell 1996]. It seems ironic that Klein had written much earlier about both modular functions and "noneuclidean geometry". He had all the expertise to make the connection, but somehow he did not. Perhaps it is understandable that he was at times less than generous to the youthful Poincaré, who had burst like a supernova on the mathematical scene.

As we have seen, Klein had plenty of reason to feel good about himself, and it would have cost nothing to be more cordial.

5. From 1893 to 1960

Between Hurwitz's paper and about 1960, there was a certain amount of routine work on automorphisms of Riemann surfaces, but very little of real significance. Fricke discovered the Hurwitz group $\mathrm{PSL}(2, 2^3)$ of order 504 and genus 7, and Wiman [1895b; 1895a] improved the bound for the order of an automorphism from $10(g-1)$ to $2(2g+1)$, which is best possible. Wiman also worked out all interesting automorphism groups for surfaces of genus 2, 3, 4, 5 and 6 by using methods of classical algebraic geometry. A lot of labour was involved.

Siegel's "remark", taking only two lines of text, was also a major contribution. One normally thinks of Siegel as an analyst and a number theorist, but here we see geometric inspiration as well.

6. Hurwitz Groups

In 1900 the most up-to-date list of finite simple groups was to be found at the end of Dickson's book [1900, Chapter XV, particularly § 290]. By 1954, there was still no advance. Then Chevalley's paper [1955] appeared, starting off the avalanche that culminated in the classification of all finite simple groups in the 1980's.

Now the search for Hurwitz groups requires a knowledge of finite simple groups. It is easy to show that the factor group of a Hurwitz group modulo any maximal normal subgroup is a simple group and also a Hurwitz group. So our obvious strategy for finding Hurwitz groups is first to comb the simple groups for Hurwitz groups and then to find extensions building on these as factor groups. See [Macbeath 1990].

Not knowing Miller's work [1902], I started from scratch. I did not need more than Dickson's book and a little basic topology of surfaces to find the following two results [Macbeath 1961; 1969].

- $\mathrm{PSL}(2,q)$, where $q=p^m$, p prime, is a Hurwitz group if and only if *either* $q=7$, *or* $q=p\equiv\pm1 \pmod 7$, *or* $q=p^3$, $p\equiv\pm2,\pm3 \pmod 7$.
- If G is a Hurwitz group of order $84(g-1)$, for $0<n\in\mathbb{Z}$, then there is a group $G(n)$ of order $84(g-1)n^{2g}$ that is also a Hurwitz group. The group $G(n)$ is an extension of a product of $2g$ copies of the finite cyclic group $\mathbb{Z}/n$ by G.

The first of these results is proved by manipulating 2×2 matrix equations in the finite field $\mathrm{GF}(q)$. The second is proved by applying Fuchsian group theory to the groups of the coarse and fine tilings just mentioned. The group of the coarse tiling is the fundamental group of the surface of genus g, which abelianizes to give a product of $2g$ copies of $\mathbb{Z}$, the infinite cyclic group. Hence the exponent $2g$ in the expression for the order of $G(n)$.

It struck me forcibly at the time, and still seems remarkable, that this is all so heavily group-theoretic. The methods indicated allow us to construct a great variety of Hurwitz groups. The second theorem allows us to derive "towers" $G(p)$, $G(p)(q)$, $G(p)(q)(r)$, $\dots$.

Indeed there is no need to look for abelian kernels only in this process. It has been observed by J. M. Cohen (oral communication) that a similar method proves that, given any simple group H, one can construct a Hurwitz group with H in its composition series. When we build towers, the order of the group, and therefore the genus of the surface, increases dramatically as a function of the number of building blocks in the tower.

In the opposite direction, Cohen [1981] proved that $\mathrm{PSL}(3,q)$ is *not* a Hurwitz group unless $q=2$.

One can also look for permutation solutions of (3). Experimentation with permutations of given degree n can be done graphically. A pair of permutations t, u satisfying (3) can be drawn as a graph with triangles for the 3-cycles of u and edges of a different colour joining points to their t-images. It is not difficult to manipulate things so that $(tu)^7=1$. The generated group is transitive if the graph is connected.

One finds by trial that the only permutation solutions for degrees 7 and 8 give us back $\mathrm{PSL}(2,7)$ — the original Klein surface! At degree 9 we find $\mathrm{PSL}(2,2^3)$, already found by Fricke. There is no transitive group for $10 \le n \le 13$. For $n=14$, we have $\mathrm{PSL}(2,13)$, and for $n=15$ we have the alternating group A_{15}.

By systematically combining graphs — and a lot of ingenuity — Conder [1980] proved that A_n is a Hurwitz group for $n \ge 168$. He also determined specifically which A_n are not Hurwitz groups for $16 \le n \le 167$.

Sah [1969] produced a lot of information about Hurwitz groups and also about other groups acting on Riemann surfaces. He showed that the Ree groups ${}^2G_2(3^p)$ are all Hurwitz groups.

As far as I know, $\mathrm{PSL}(2,q)$, $\mathrm{PSL}(3,q)$, A_n, and the Ree groups ${}^2G_2(3^p)$ are the only infinite series of finite simple groups where we know precisely which ones are Hurwitz groups.

During the search for finite simple groups, eleven “sporadic” simple groups were found to be Hurwitz groups. These are listed in Conder’s excellent survey article [1990]. It contains some more techniques for producing Hurwitz groups, and is the best place to get further information about them.

7. The Wider Picture

The Hurwitz groups, then, proved to be surprisingly interesting. Apart from Miller’s paper, presentations including (3) are found scattered through the literature. In [Coxeter and Moser 1957, p. 96] we find a presentation displaying $\mathrm{PSL}(2,7)$ as a Hurwitz group and $\mathrm{PSL}(2,13)$ as a Hurwitz group in two different ways. We can deduce that $\mathrm{PSL}(2,13)$ acts on *two* Riemann surfaces of genus 14. (We know now that there is a third one.)

Now, the first few Hurwitz groups, in order, act on surfaces of genus 3, 7, 14 and 17, and the admissible genera seem to become more sparse as they get larger. For every $g \ge 2$, though, there is a maximum order $\mu(g)$ for an automorphism group, and it is not difficult to show that, for any $g \ge 2$ there is a group of order

$8g+8$ acting on some surface of genus g. We therefore have

$$8g+8 \leq \mu(g) \leq 84(g-1). \tag{4}$$

Klein's and other surfaces show that the upper bound is sharp. Independently, and about the same time, Bob Accola in Providence and Colin Maclachlan in Birmingham, England, found the lower bound and proved that it too is sharp. Each of them produced an infinite family of g with $\mu(g) = 8g+8$, and the two families are not only distinct but disjoint! Maclachlan used the language of Fuchsian groups, but Accola, like Hurwitz, worked without them. See [Accola 1968; Maclachlan 1969].

Folk also looked at special kinds of groups. Wiman had dealt with *cyclic* groups, but there was more to say. Harvey [1966] found, for each n, the smallest genus of a surface with an automorphism of order n. This gave Wiman's bound as an easy corollary. Maclachlan [1965] found the upper bound for *abelian* automorphism groups. Accola was the first to observe that the order of a *soluble* group of automorphisms cannot exceed $48(g-1)$. Zomorrodian [1985] found the upper bound for *nilpotent* groups. See also [Macbeath 1984]. Maclachlan and Gromadzki [1989] found the bound for *supersoluble* groups.

The aim of these workers was to find the largest or smallest group in some particular category, but there is a good reason for looking at *all* the automorphism groups acting on surfaces of a given genus, whether or not they have any extreme value. Here is why.

For every g we have a *Teichmüller space* $\mathbb{T}_g$ of "marked Riemann surfaces" analogous to the modular figure in genus 1. Topologically $\mathbb{T}_g$ is a euclidean space of $6g-6$ real dimensions. The *mapping class group* M_g is a discontinuous group acting on $\mathbb{T}_g$. The quotient space $\mathbb{T}_g/M_g$ is the space R_g of all closed Riemann surfaces of genus g. The quotient mapping $\mathbb{T}_g \to R_g$ is a branched covering and the points where ramification occurs are the Riemann surfaces with nontrivial automorphisms.

To understand this situation it is necessary to get some understanding of the whole set of groups involved as well as the dimensions of the subspaces of $\mathbb{T}_g$ consisting of the fixed point sets for each group.

Though we have a good understanding of Teichmüller space, there is a lot we don't know about the configuration of interlocking fixed point sets, or *branch loci*, as they are called. Even for fairly small values of g, the number of possible groups, including cyclic and dihedral groups, is quite large and the same group may act in several topologically different ways, as we saw with $\mathrm{PSL}(2,13)$.

Some people have tried to outflank the problem, by taking a given group and finding all the genera of surfaces on which it acts. Harvey [1966] did this for cyclic groups, and his work was extended by Lloyd [1972]. More recently, Kulkarni [1987] has shown that, for *any* group, the admissible values of g settle ultimately into a periodic pattern modulo the prime factors of its order. For more information see [Kulkarni 1987; Harvey 1971; Macbeath and Singerman 1975]. Much remains to be done in this direction.

The title of Hurwitz's paper, freely translated, is *Algebraic structures with one-to-one self-mappings*. Most of the paper deals with the general picture of an algebraic curve with automorphisms, looking closely at the branched covering of the quotient surface by the target surface. The bound $84(g-1)$ falls out as a by-product.

The spirit of Hurwitz's work is consistent with studying the general picture and not beoming obsessed with one particular Fuchsian group, which, by an arithmetical accident, happens to have the smallest quotient area. That is in the spirit of Klein and Fricke too. Their books on modular and automorphic functions [Klein and Fricke 1890–92; Fricke and Klein 1912] give many examples of curves with fairly large non-Hurwitz automorphism groups.

8. Conclusion

It is appropriate to reflect how much Klein knew about his curve, and how little we know about all the Hurwitz surfaces we have constructed. Apart from Klein's curve and the curve of genus 7, we know equations for no other curve with $84(g-1)$ automorphisms. Each one of them is an isolated point of $\mathbb{T}_g$, so the problem makes good sense. The only really useful tool that has emerged in looking for equations seems to be the Lefschetz fixed point formula [Macbeath 1965; 1973]. For $\mathrm{PSL}(2,2^3)$, it worked, but for $\mathrm{PSL}(2,13)$ it was not quite enough. On the other hand, there are limits to what we can expect to do. The genus of the curve on which the Hurwitz group A_{15} acts is 7783776001. Even with modern computers, the calculation of an equation might be difficult even if we had a program to do it.

Let us pay tribute to Klein: he may not have known as much group theory as we do, but he knew a whole lot more about other things.

References

[Accola 1968] R. D. M. Accola, "On the number of automorphisms of a closed Riemann surface", *Trans. Amer. Math. Soc.* **131** (1968), 398–408.

[Chevalley 1955] C. Chevalley, "Sur certains groupes simples", *Tōhoku Math. J.* (2) **7** (1955), 14–66.

[Cohen 1981] J. Cohen, "On non-Hurwitz groups and noncongruence subgroups of the modular group", *Glasgow Math. J.* **22**:1 (1981), 1–7.

[Conder 1980] M. D. E. Conder, "Generators for alternating and symmetric groups", *J. London Math. Soc.* (2) **22**:1 (1980), 75–86.

[Conder 1990] M. Conder, "Hurwitz groups: a brief survey", *Bull. Amer. Math. Soc. (N.S.)* **23**:2 (1990), 359–370.

[Coxeter and Moser 1957] H. S. M. Coxeter and W. O. J. Moser, *Generators and relations for discrete groups*, Springer, Berlin, 1957. Fourth edition, 1980.

[Dickson 1900] L. E. Dickson, *Linear Groups, with an exposition of the Galois Field Theory*, Chicago, Leipzig, 1900. Reprinted by Dover, New York, 1958.

[Fricke 1926] R. Fricke, *Lehrbuch der Algebra*, Vieweg, Braunschweig, 1926.

[Fricke and Klein 1912] R. Fricke and F. Klein, *Vorlesungen über die theorie der automorphen Functionen, 2: die functionentheoretischen Ausführungen und die Anwendungen*, Teubner, Leipzig, 1912. Reprinted by Johnson Reprint, New York, and Teubner, Stuttgart, 1965.

[Gromadzki and Maclachlan 1989] G. Gromadzki and C. Maclachlan, "Supersoluble groups of automorphisms of compact Riemann surfaces", *Glasgow Math. J.* **31**:3 (1989), 321–327.

[Harvey 1966] W. J. Harvey, "Cyclic groups of automorphisms of a compact Riemann surface", *Quart. J. Math. Oxford* (2) **17** (1966), 86–97.

[Harvey 1971] W. J. Harvey, "On branch loci in Teichmüller space", *Trans. Amer. Math. Soc.* **153** (1971), 387–399.

[Hurwitz 1893] A. Hurwitz, "Über algebraische Gebilde mit eindeutigen Transformationen in sich", *Math. Annalen* **41** (1893), 403–442.

[Jordan 1878] M. C. Jordan, "Mémoire sur les équations différentielles linéaires à intégrale algébrique", *J. für reine angew. Math.* **84** (1878), 89–215. Reprinted as pp. 13–140 of his *Œuvres*, vol. 2, Gauthier-Villars, Paris, 1962.

[Klein 1879] F. Klein, "Ueber die Transformationen siebenter Ordnung der elliptischen Funktionen", *Math. Annalen* **14** (1879), 428–471. Reprinted as [Klein 1923, LXXXIV, pp. 90–136]. Translated in this collection.

[Klein 1923] F. Klein, *Gesammelte Mathematische Abhandlungen, 3: Elliptische Funktionen* etc., edited by R. Fricke et al., Springer, Berlin, 1923. Reprinted by Springer, 1973.

[Klein and Fricke 1890–92] F. Klein, *Vorlesungen über die Theorie der elliptischen Modulfunctionen,* ausgearbeitet und vervollständigt von Robert Fricke (2 vol.), Teubner, Leipzig, 1890–92. Reprinted by Johnson Reprint, New York, 1966.

[Kulkarni 1987] R. S. Kulkarni, "Symmetries of surfaces", *Topology* **26**:2 (1987), 195–203.

[Lloyd 1972] E. K. Lloyd, "Riemann surface transformation groups", *J. Combinatorial Theory Ser. A* **13** (1972), 17–27.

[Macbeath 1961] A. M. Macbeath, "On a theorem of Hurwitz", *Proc. Glasgow Math. Assoc.* **5** (1961), 90–96.

[Macbeath 1965] A. M. Macbeath, "On a curve of genus 7", *Proc. London Math. Soc.* (3) **15** (1965), 527–542.

[Macbeath 1969] A. M. Macbeath, "Generators of the linear fractional groups", pp. 14–32 in *Number Theory* (Houston, 1967), edited by W. J. Leveque and E. G. Straus, Proc. Sympos. Pure Math. **12**, Amer. Math. Soc., Providence, 1969.

[Macbeath 1973] A. M. Macbeath, "Action of automorphisms of a compact Riemann surface on the first homology group", *Bull. London Math. Soc.* **5** (1973), 103–108.

[Macbeath 1984] A. M. Macbeath, "Residual nilpotency of Fuchsian groups", *Illinois J. Math.* **28**:2 (1984), 299–311.

[Macbeath 1990] A. M. Macbeath, "Automorphisms of Riemann surfaces", pp. 107–112 in *Combinatorial group theory* (College Park, MD, 1988), edited by B. Fine et al., Contemporary Mathematics **109**, Amer. Math. Soc., Providence, RI, 1990.

[Macbeath and Singerman 1975] A. M. Macbeath and D. Singerman, "Spaces of subgroups and Teichmüller space", *Proc. London Math. Soc.* (3) **31**:2 (1975), 211–256.

[Maclachlan 1965] C. Maclachlan, "Abelian groups of automorphisms of compact Riemann surfaces", *Proc. London Math. Soc.* (3) **15** (1965), 699–712.

[Maclachlan 1969] C. Maclachlan, "A bound for the number of automorphisms of a compact Riemann surface", *J. London Math. Soc.* **44** (1969), 265–272.

[Miller 1902] G. A. Miller, "Groups defined by the order of two generators and the order of their product", *Amer. J. Math.* **24** (1902), 96–100.

[Poincaré 1882] H. Poincaré, "Théorie des groupes fuchsiens", *Acta Mathematica* **1** (1882), 1–62.

[Sah 1969] C.-h. Sah, "Groups related to compact Riemann surfaces", *Acta Math.* **123** (1969), 13–42.

[Siegel 1945] C. L. Siegel, "Some remarks on discontinuous groups", *Ann. of Math.* (2) **46** (1945), 708–718.

[Stillwell 1996] J. Stillwell, *Sources of hyperbolic geometry*, History of Mathematics, Amer. Math. Soc. and London Math. Soc., 1996.

[Wiman 1895a] A. Wiman, "Über die hyperelliptischen Curven und diejenigen von Geschlechte $p = 3$, welche eindeutigen Transformationen in sich zulassen", *Bihang till Kongl. Svenska vetenskaps-akademiens handlingar* **21**:I-1 (1895), 1–23.

[Wiman 1895b] A. Wiman, "Ueber die algebraischen Curven von den Geschlechten $p = 4$, 5 und 6, welche eindeutigen Transformationen in sich besitzen", *Bihang till Kongl. Svenska vetenskaps-akademiens handlingar* **21**:I-3 (1895), 1–41.

[Zomorrodian 1985] R. Zomorrodian, "Nilpotent automorphism groups of Riemann surfaces", *Trans. Amer. Math. Soc.* **288**:1 (1985), 241–255.

A. Murray Macbeath
1 Church Hill Court
Lighthorne
Warwick CV 35 0AR
England
murray@maths.warwick.ac.uk

The Eightfold Way
MSRI Publications
Volume **35**, 1998

From the History of a Simple Group

JEREMY GRAY

The attractive pattern of 168 shaded and 168 unshaded triangles shown in Figure 1 has an interesting history. Since its discovery by Klein in 1878 (see

Figure 1

This article originally appeared in *The Mathematical Intelligencer* **4**:2 (1982), 59–67. The editor is thankful to the author and to Springer-Verlag for permission to reprint.

[Klein 1879]), it has often been reproduced; a close cousin (Figure 2) inspired the badge of the 1978 International Congress of Mathematicians in Helsinki. This article considers its origins, which lie in the fields of nineteenth century geometry and the theory of equations.

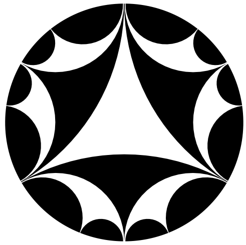


Figure 2

But first let us look closely at the figure itself. Each triangle, shaded or unshaded, has angles of $\pi/2$, $\pi/3$, and $\pi/7$. Since

$$\frac{\pi}{2}+\frac{\pi}{3}+\frac{\pi}{7}=\frac{41\pi}{42}<\pi,$$

we immediately recognize that this is a non-Euclidean figure, but we shall see that Klein missed this conclusion altogether.

In each of the 14 slices emanating from the center there are 12 shaded and 12 unshaded triangles, so there are 168 of each kind. The sides of each triangle are arcs of circles orthogonal to the boundary circles, or are diameters. The figure can be continued in this fashion to reach indefinitely close to the boundary, and it provides in this way a tessellation of the non-Euclidean plane. The unshaded tessellation is preserved by non-Euclidean reflection in any side of any triangle (i.e., by inversion) and so has the group of all such reflections as its symmetry group. The group generated by all products of pairs of reflections is the symmetry group of the shaded figure.

Klein had been led to construct the figure because of its use in studying a certain polynomial equation (described at the end of this paper) for which the group permuting the roots is $\mathrm{PSL}(2;\mathbb{Z}/7\mathbb{Z})$, sometimes known as G_{168} because of the number of its elements. Our first task, then, is to understand this group geometrically.

The map $\mathbb{Z}\to\mathbb{Z}/7\mathbb{Z}$ which takes residues modulo 7 induces a homomorphism between two groups of 2×2 matrices:

$$\mathrm{SL}(2;\mathbb{Z})\to\mathrm{SL}(2;\mathbb{Z}/7\mathbb{Z}),$$

where $\mathrm{SL}(2,K)$ is the group of 2×2 matrices with entries in K and of determinant 1. This is an onto map, and we shall denote its kernel by Γ_7. The group $\mathrm{SL}(2;\mathbb{Z})$ acts on the upper half-plane $H=\{z\in\mathbb{C}:\mathrm{Im}(z)>0\}$: the element $\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)$ in $\mathrm{SL}(2;\mathbb{Z})$ sends z to $(az+b)/(cz+d)$.

Since $\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}-a&-b\\-c&-d\end{smallmatrix}\right)$ have the same effect on all $z\in H$, it is sometimes convenient to factor out the centre, $\{\pm 1\}$, of $\mathrm{SL}(2;\mathbb{Z})$, and obtain $\mathrm{PSL}(2;\mathbb{Z})=\mathrm{SL}(2;\mathbb{Z})/\{\pm 1\}$, which acts faithfully on H. Dedekind was the first to describe this group geometrically, in a very important paper [1878]. He defined the region

$$R=\left\{z\in H:|z|\geq 1,\ -\tfrac{1}{2}\leq\mathrm{Re}(z)\leq\tfrac{1}{2}\right\}$$

(see Figure 3), and showed that the orbit

$$O(z)=\left\{\frac{az+b}{cz+d}:\begin{pmatrix}a&b\\c&d\end{pmatrix}\in\mathrm{SL}(2;\mathbb{Z})\right\}$$

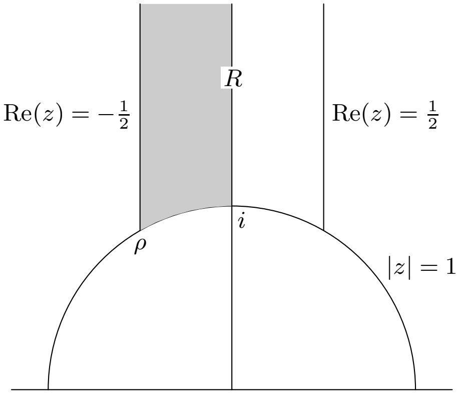


Figure 3

of each point z of H meets R precisely once (in its interior) or twice (on its boundary). Consequently $\mathrm{SL}(2;\mathbb{Z})$ moves this region around en bloc, and covers H like a tile, with overlaps occurring only on copies of the boundary of R. Moreover, as Dedekind said, the elements of $\mathrm{SL}(2;\mathbb{Z})$ are all products of these matrices:

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \text{which sends } z \text{ to } -1/z \text{ and fixes } i, \text{ and}$$

$$\begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}, \quad \text{which sends } z \text{ to } -1/(z+1) \text{ and fixes } \rho,$$

a cube root of unity.

Klein took over Dedekind's simple geometric presentation, and refined it by making explicit that the set of all 2×2 matrices with integer entries and determinant 1 is a group, a fact Dedekind had not stressed although he would have been well aware of it, and by looking for particular subgroups of it. The one of most interest to him was Γ_7.

The index of Γ_7 in $\mathrm{SL}(2;\mathbb{Z})$ is, of course, the order of $\mathrm{SL}(2;\mathbb{Z}/7\mathbb{Z})$. Since Galois's work had been published (in 1846) it had been usual to consider the action of this group on the eight symbols $0, 1, \ldots, 7, \infty$ by

$$z \mapsto \frac{\alpha z + \beta}{\gamma z + \delta}, \quad \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}(2;\mathbb{Z}/7\mathbb{Z}).$$

These symbols can be regarded as the slopes of lines through the origin in the plane defined over the field of 7 elements (more precisely, as the points of the projective line over that field). So the group $\mathrm{SL}(2;\mathbb{Z}/7\mathbb{Z})$ has 336 elements, for there are 8 directions for the position of the first basis vector $\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ under $\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$, each with 6 possible positions for the image of $\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ itself, then there are 7 choices for the direction of the image of $\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$, but no choice for its position once the direction is chosen, since $\alpha\delta - \beta\gamma = 1$: $8 \cdot 6 \cdot 7 = 336$. The group $\mathrm{PSL}(2;\mathbb{Z}/7\mathbb{Z})$ therefore has $\frac{1}{2}336 = 168$ elements. So, looking at the faithful action, one finds that $\bar{\Gamma}_7 = \Gamma_7/\{\pm 1\}$ has index 168 in $\mathrm{PSL}(2;\mathbb{Z})$. So it must move 168 copies of R around en bloc, and a suitable choice of which 168 copies can be made depending on the purposes at hand. One way is to observe that

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix},$$

which is equivalent to $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ in $\mathrm{PSL}(2;\mathbb{Z})$, an element which has the effect of shifting the triangle sideways by $z \mapsto z+1$. Since

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix},$$

the element

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^7$$

is in $\bar{\Gamma}_7$, so one picks 7 copies of R in a horizontal strip. One can also pick other elements of $\bar{\Gamma}_7$ which yield other copies of R, until the 168 block is determined.

To study the quotient, G_{168}, one observes that any matrix representative of an element in it also moves the 168-member block around, but that action is only defined modulo $\bar{\Gamma}_7$, so G_{168} really maps the 168-member block to itself, once suitable identifications have been made. The case G_{168} is rather unwieldy at first glance, so consider for a moment starting with residues modulo 2:

$$\bar{\Gamma}_2 \to \mathrm{PSL}(2;\mathbb{Z}) \to \mathrm{PSL}(2;\mathbb{Z}/2\mathbb{Z}).$$

$\mathrm{PSL}(2;\mathbb{Z}/2\mathbb{Z})$ has only 6 elements:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

So $\bar{\Gamma}_2$ moves 6 copies of R around en bloc, and $\mathrm{PSL}(2;\mathbb{Z}/2\mathbb{Z})$ can be regarded as a group of self-maps of that region. The most attractive picture of this is shown in Figure 4, and one notices that the action of $\bar{\Gamma}_2$, unlike that of $\mathrm{PSL}(2;\mathbb{Z})$, is fixed-point free. It is this figure that inspired the ICM badge.

The action of $\bar{\Gamma}_2$ identifies the edges of the larger region in pairs, and so one can ask what the region is topologically. In this case it is clear from the identifications that the region is a sphere.

There is one problem with these pictures: the vertex at ∞ of the region R. Klein simply switched to a region $\tilde{R}$ where this angle was $2\pi/7$, since $\bar{\Gamma}_7$ cycles 7 copies of R around the vertex. This can be done by standard moves in the theory of complex functions: either appeal to the Riemann mapping theorem, or find an explicit map of R holomorphic everywhere except at copies of ∞, where it has a suitable branch point. Klein presumably did the first; subsequently a student of his, the American mathematician M. W. Haskell, did the second using a quotient of two solutions to a hypergeometric equation [Haskell 1891]. Finally we have Figure 1 before us, together with a description of G_{168} as the self-maps

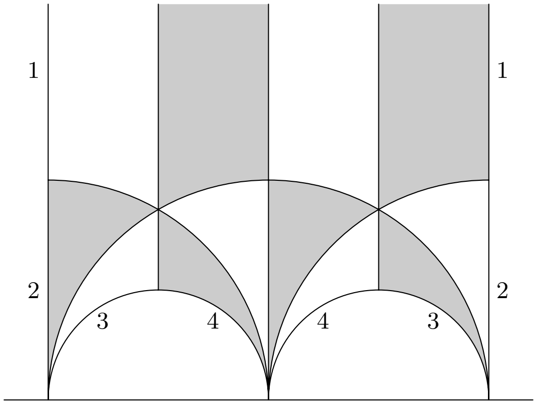


Figure 4

of this region, thought of as the quotient space $\bar{\Gamma}_7\backslash H$, which also preserve the shading. It turns out that the edge identifications are 1 with 6, 3 with 8 and so on, and the even- and odd-numbered edges are directed oppositely so that the triangles match up.

Klein was interested in the figure as an algebraic curve and as a Riemann surface so he wanted to know its genus. This can be found using Euler's formula

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

where V, E, and F denote the number of vertices, edges and faces in a triangulation of the surface, and g is the genus.

Happily, we have a triangulation: it has 336 triangles (since $\bar{R}$ is made up of 2 triangles), so $F = 336$; and $E = 336 \cdot 3/2$, since each edge is counted twice. As for the vertices, 336 have angles of $\pi/7$, so 14 cluster together at each one; 336 have angles of $\pi/3$ and are identified in 6's; and 336 have angles of $\pi/2$ and are identified in 4's: a total of

$$336\left(\frac{1}{14} + \frac{1}{6} + \frac{1}{4}\right) = 164,$$

so $V - E + F = 4$, and $g = 3$: the Riemann surface has genus 3. Klein followed Riemann's approach of looking at the order of the branch points in order to calculate the genus.

It then follows from Riemann's work on algebraic curves [1857, § 13] that the surface must be describable as a quartic, that is, using homogeneous coordinates $[x, y, z]$, by a homogeneous polynomial of degree 4. Klein then turned to the projective theory of higher plane curves that had been developed in the preceding generation, and showed how it could be illuminated by his new methods of Riemann surface theory. This is a path well worth following.

Plane Algebraic Curves

Although Newton had provided a very thorough analysis of cubics in the 1670's, mathematicians rather neglected the systematic study of algebraic curves other than conics until the start of the nineteenth century. When finally they began, they confronted the question of deciding what could interestingly be said about the profusion of new cubics, quartics, quintics, and so forth with which they were confronted. The properties that most attracted them were projective in nature, and were not shared by conics, notably: points of inflection, bitangents (lines tangent at two places to a curve), double points, and cusps. The pioneer in this study was Plücker, who was Klein's first mathematical teacher. Plücker [1835] showed that a non-singular curve $F(x, y, z) = 0$ of degree n has $3n(n-2)$ inflection points. Hesse's proof [1844] is simpler, being couched in homogeneous coordinates, but it essentially followed Plücker's argument. Hesse observed that at an inflection point adjacent normals are parallel, and so the mean curvature vanishes there. But the formula for the mean curvature is

$$\left|\frac{\partial^2 F}{\partial x_i \partial x_j}\right|,$$

which equated to zero is a curve of degree $3(n-2)$. So, by Bezout's principle, it meets $F = 0$ in $3(n-2)n$ points, which are the points of inflection.

This result led Plücker to make an intriguing observation in his next book [1839]. The tangent to $F = 0$ at p has equation

$$x_1 \frac{\partial F}{\partial x_1}(p) + x_2 \frac{\partial F}{\partial x_2}(p) + x_3 \frac{\partial F}{\partial x_3}(p) = 0$$

and the triple

$$\left(\frac{\partial F}{\partial x_1}(p), \frac{\partial F}{\partial x_2}(p), \frac{\partial F}{\partial x_3}(p)\right)$$

can be thought of as defining the line coordinates of the tangent. This triple can be thought of as a point in the dual space to the original projective plane, and thus as defining a new plane curve called the dual of the original curve (see Figure 5). Geometrically this can be done by picking a circle and then replacing each point to the original curve by its polar with respect to the circle, and looking at the envelope of the polars. Both methods were used. What is the degree of the dual curve? Poncelet [1832] had shown that the tangents through (ξ_1, ξ_2, ξ_3) to $F = 0$ had equations

$$\xi_1 \frac{\partial F}{\partial x_1}(p) + \xi_2 \frac{\partial F}{\partial x_2}(p) + \xi_3 \frac{\partial F}{\partial x_3}(p) = 0$$

for suitable P on $F = 0$. The locus of all points in the plane for which this equation is true (for a given $[\xi_1, \xi_2, \xi_3]$ and F) is a curve of degree $n - 1$ called a first polar of F. It meets $F = 0$ in $n(n-1)$ points, so in general there are $n(n-1)$ tangents to a given curve of degree n from a given point. Consequently

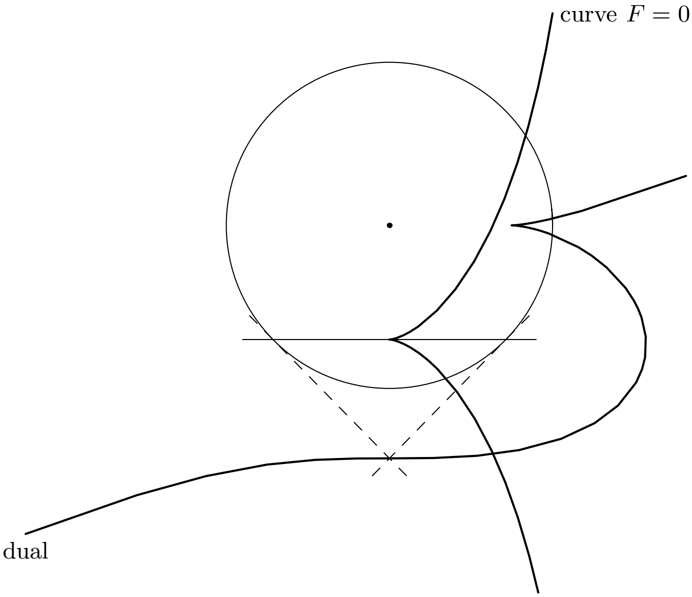


Figure 5

the dual curve to a curve of degree n is of degree $n(n-1)$, for the dual of $n(n-1)$ lines through a point is $n(n-1)$ points on a line.

Plücker's intriguing paradox is this: plainly the dual curve of a dual curve is the original curve, but the degree formula shows that the double dual has degree $n(n-1)(n(n-1)-1)$, which is greater than n as soon as $n > 2$. Plücker had a solution: any line through a double point is a tangent, since it meets the curve in two coincident points, so the first polar passes through that point. But this is not what tangency is really about, and if those intersections are ignored by pulling the double point apart, this means two intersections of a curve and its first polar must be discussed. So each double point on the original curve lowers the degree of its dual by 2. Moreover, if the curve has a cusp the first polar is a tangent there, so each cusp lowers the degree by 3. For example, the curve $x_1^2 x_3 - x_2^3 = 0$ has a cusp at $[0,0,1]$. Its first polar with respect to $[0,0,1]$ is $x_1^2 = 0$, which in fact is the equation of the tangent in this example.

So, if the original curve has α bitangents and β inflection points, the dual will have α double points and β cusps, since bitangents dualize to double points and inflection points to cusps. So if

$$2\alpha + 3\beta = n(n-1)\bigl(n(n-1)-1\bigr) - n = n^3(n-2),$$

the paradox is explained. Moreover, Plücker had already shown that $\beta = 3n \times (n-2)$, from which he deduced that

$$\alpha = \tfrac{1}{2}n(n-2)(n^2-9),$$

and he proclaimed that a nonsingular curve of degree n has a dual with α double points, β cusps, and degree $n(n-1)$, where

$$2\alpha + 3\beta = n^3(n-2).$$

This formula is now called Plücker's formula. The value of α was first calculated directly by Jacobi [1850]. The particular case when $n = 4$ is of most interest to us: a non-singular quartic should have 24 inflection points and 28 bitangents. The inflection points cannot all be real, but the bitangents can be, and Plücker even gave an example [1839]. He took two degenerate quartics: the four straight lines in Figure 6, and the circle counted twice. A linear combination of the equations of these two curves defines a quartic with double points at $(0,0)$, $(1,1)$, $(1,-1)$, and a vertical tangent at the points a, a'. The figure shows in thin lines the particular combination

$$\Omega_4 = (y + xy)(y - x)(x - 1)(x - 1.85) - 5(y^2 + x(x-2))^2 = 0.$$

When deformed into $\Omega_4 - k = 0$, for $k > 0$ small, the curve splits apart into four bean-shaped pieces — the thick curves of the figure. Each has one bitangent of its own and each pair has 4, a total of $4 + 6 \cdot 4 = 28$ in all. (We have varied

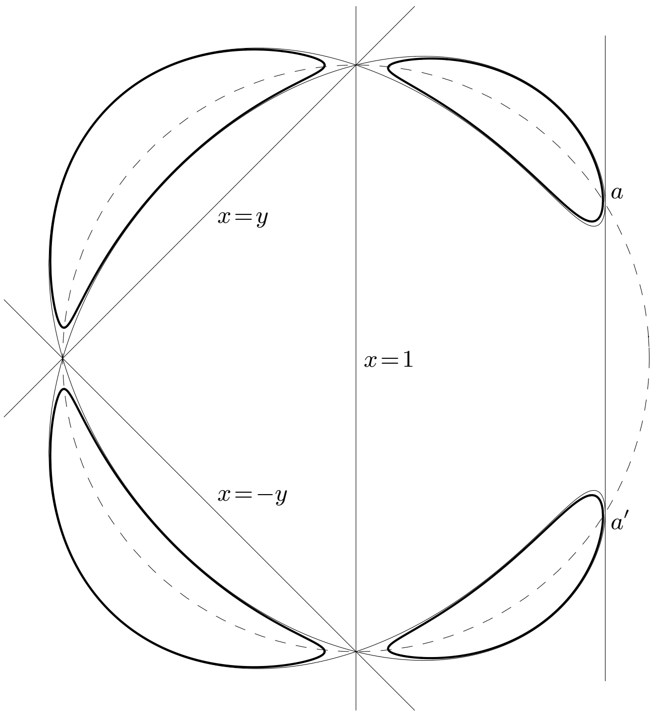


Figure 6

Plücker's coefficients, which were $\frac{3}{2}$ and 2 instead of 1.85 and 5, in order to make the concavity more apparent.)

The 28 bitangents became, and remain, a topic of delight. They are, for instance, intimately connected to the 27 lines on a cubic surface, a fact first noticed by Geiser [1869], and their automorphism group is isomorphic to the Weyl group of the exceptional Lie algebra E_7. Their history is far too lengthy to describe here, even in the period before Klein, but mention should be made of Hesse's paper [1855], in which he studied them via the 28 lines through 8 points in space, and in particular, to work of Riemann.

Riemann's "Theorie der Abelschen Functionen" introduced an (infinitely many valued) "function" θ of p variables on a Riemann surface of genus p which was crucial to his solution to the Jacobi inversion problem for integrals. He associated what he called a characteristic to θ, an expression

$$\begin{pmatrix} e_1, e_2, \ldots, e_p \\ e'_1, e'_2, \ldots, e'_p \end{pmatrix},$$

where each entry is 0 or 1, and he said the characteristic was odd if

$$e_1 e'_1 + e_2 e'_2 + \cdots + e_p e'_p$$

was odd, and even otherwise. Induction on p shows that $(2^{p-1})(2^p - 1)$ characteristics are odd.

When the characteristic is odd θ has two repeated zeros and 2 repeated poles on the surface, so it could be made to yield a bitangent curve to the surface, and when $p = 3$ indeed to yield a bitangent. All this material, although partly published by Riemann, his student Roch, and by Clebsch, was very obscure to Riemann's contemporaries. Clebsch himself thought this was due to the elusive nature of the θ-function, which was defined transcendentally and only after a long series of boldly innovative remarks. Riemann's paper defines Riemann surfaces and studies them topologically, uses the contentious Dirichlet principle to prove an index theorem for the genus, considers what functions can exist on a Riemann surface and proves the Riemann inequality for the dimension of the space of meromorphic functions with prescribed poles, discusses coordinate transformations and birational transformations of a given curve and the dimension of the corresponding moduli space of inequivalent curves of a given genus, and proves half of Abel's theorem before getting round to Jacobi inversion. Little wonder people found it difficult! But the notation of the characteristics was convenient, and in 1874 Weber used it to describe how the 28 bitangents are related (see [Weber 1876]).

Briefly, Steiner had shown in [1848] that the bitangents fell into 63 groupings of 6 pairs, with the property that the contact points of each pair with the quartic gave a set of 8 points lying on a conic. Weber showed that the 63 families could be indexed by the 63 characteristics other than $\binom{0\,0\,0}{0\,0\,0}$ and that the sum of each pair of characteristics in a grouping was the indexing characteristic. Thus $\binom{0\,0\,1}{0\,0\,0}$ indexes the pairs

$$\binom{1\,0\,1}{1\,0\,0},\ \binom{1\,0\,0}{1\,0\,0};\quad \binom{0\,1\,1}{0\,1\,0},\ \binom{0\,1\,0}{0\,1\,0};$$
$$\binom{1\,1\,1}{1\,0\,0},\ \binom{1\,1\,0}{1\,0\,0};\quad \binom{1\,1\,1}{0\,1\,0},\ \binom{1\,1\,0}{0\,1\,0};$$
$$\binom{0\,1\,1}{1\,1\,0},\ \binom{0\,1\,0}{1\,1\,0};\quad \binom{1\,0\,1}{1\,1\,0},\ \binom{1\,0\,0}{1\,1\,0}$$

and

$$\binom{1\,0\,1}{1\,0\,0} + \binom{1\,0\,0}{1\,0\,0} = \binom{0\,0\,1}{0\,0\,0},$$

etc. (adding coordinatewise, mod 2). This approach was in fact that of Riemann [1862], as Weber found out on becoming joint editor of Riemann's *Werke* with Dedekind later in 1874, and had earlier been taken by Clebsch [1864], which Weber seems not to have known. The geometric situation is that two conics, each touching a quartic in 4 points, lie in the same system if their 8 points lie in a conic. There are 63 systems (a result of Hesse's) and each system contains 6 line pairs, the pairs of bitangents.

Jordan, basing himself on Clebsch's work — which was published, rather than on Riemann's, which was not — gave an analysis of the 28 bitangents in [Jordan 1870]. He showed (§ 332) that the group of symmetries of the bitangents is isomorphic to the symplectic group $\mathrm{Sp}(6;\mathbb{Z}/2\mathbb{Z})$, that is, to the group of 6×6 matrices over the field of 2 elements which preserves the inner product represented by the matrix

$$A = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \equiv \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \pmod 2,$$

where I is the 3×3 identity matrix. He also showed (§ 455) that the subgroup of the group of symmetries which fixes a bitangent is isomorphic to the group of symmetries of the 27 lines in a cubic surface, thereby connecting his work to Geiser's.

It is clear to us that each characteristic is a vector in the six-dimensional vector space over $\mathbb{Z}/2\mathbb{Z}$ and that the bitangents corresponding to those vectors $\boldsymbol{v}$ for which $\boldsymbol{v}^T A\boldsymbol{v} \equiv 1 \pmod 2$; and, moreover, the action of G_{168} in its alternative guise as $\mathrm{SL}(3;\mathbb{Z}/2\mathbb{Z})$ is also now apparent. This version of G_{168} was presented by Weber in [1896, p. 539], where he attributed it to Kronecker. However, Jordan did not use this geometric approach, nor did Dickson in his discussion [1900]. It would be interesting to know who first interpreted the characteristics in terms of finite geometries, thereby making explicit what was implicit, but not geometric, in [Jordan 1870]. The American mathematician A. B. Coble [1908; 1913] seems to have been the first to illuminate the 27 lines and 28 bitangents with the elementary theory of geometries over finite fields.

The combinatorial aspects of all this are pleasant, but the mathematics is certainly not easy. All the more attractive then for Klein when he saw how to make some of these configurations visible in this picture of his Riemann surface.

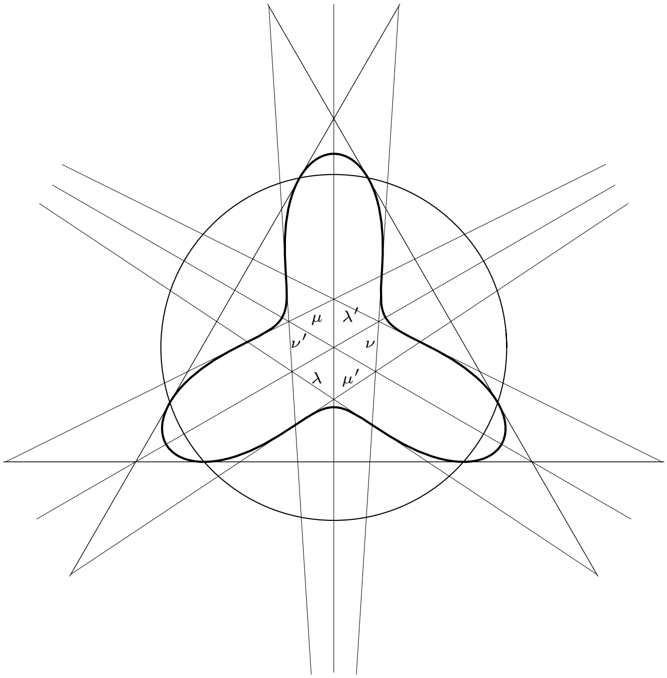


Figure 7. From [Klein 1879].

Let us return to the description of the surface as an algebraic curve. It is a quartic, and Klein showed quickly that if three suitable inflection tangents are taken as triangle of reference the equation can be written as

$$F(x_1, x_2, x_3) = x_1^3 x_2 + x_2^3 x_3 + x_3^3 x_1 = 0.$$

As a real locus it looks like the curve in Figure 7. Under the action of G_{168} a typical point has an orbit consisting of 168 points. But some have smaller orbits: the vertices have orbits consisting of 24, 56, and 84 points. These must, by simple considerations of invariant theory, said Klein, correspond to the 24 inflection points, the 56 bitangents points, and 84 sextatic points (where a conic has sixfold contact with the curve). So there are all these points, hitherto hard to visualize, all laid out in one figure. Klein called them $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ points respectively.

To get at the inter-relations of these points, Klein used elementary matrix algebra and group theory to give an exhaustive analysis of the subgroups of G_{168}. He found, amongst other subgroups, 14 of order 4 (now called Klein's group), 14 of order 24, which come in two families of 7 conjugates, and 28 non-abelian groups of order 6. The analysis showed that G_{168} has no non-trivial normal subgroups, but Klein did not remark explicitly on its simplicity. Indeed, he was more interested in the existence of 8 conjugate subgroups of order 21, and of two families of 7 conjugate subgroups of order 24 which are isomorphic to the group of proper motions of an octahedron.

It is an easy matter to carry out the matrix algebra to find all the elements of a given order. For example, elements of order 2 must have vanishing trace. One can then arrange them in conjugacy classes, and thence find all the subgroups of a given order. But it is not so easy to see them in the figure, because the action of the group is not so clear and it is difficult to work with the identifications. Klein argued as follows. The $\boldsymbol{b}$-points, for example, are each fixed by a rotation of order 3. There are 28 groups of order 3 in G_{168}, so each such group fixes a pair of $\boldsymbol{b}$-points and these are the points of contact of a bitangent. The $\boldsymbol{c}$-points are fixed by rotations of order 2, of which there are 21, so each such rotation fixes 4 of them.

To display them in the figure, Klein considered the 28 "symmetry lines" which run in a cycle through six $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$-points, as, for example, do the lines, which run straight from the central point. They must be pursued with care across the identifications, especially at the two kinds of vertex. There are 2 symmetry lines through each $\boldsymbol{c}$-point. There is then a unique pair of such lines which does not meet the first pair, and the 4 $\boldsymbol{c}$-points so picked out are a typical set of 4 $\boldsymbol{c}$-points fixed by a rotation. Similarly the 3 symmetry lines through a $\boldsymbol{b}$-point meet again in the corresponding $\boldsymbol{b}$-point. Klein denoted 4 such pairs in the figure by A, A'; B, B'; C, C'; D, D'. However, he was not able to go further with this analysis and exhibit the 63 systems of 6 pairs of bitangents. The Riemannian theory of theta-characteristics is not taken up in this paper, nor in the famous study [Klein 1882], and one rather supposes that Klein, like his contemporaries, did not really understand it.

But he did have other, new, things to say. He showed that G_{168} could be written down as a group acting on complex projective two-space, since it was the automorphism group of a plane projective curve, and he exhibited its generators explicitly. This gave him a finite subgroup of $\mathrm{SL}(3;\mathbb{C})$, and he observed that it was missing from Jordan's list of such groups [Jordan 1878], which he had published earlier while studying differential equations all of whose solutions are algebraic. Jordan, a friend of Klein's since the latter's visit to Paris in 1870, accepted the correction in a letter and in a subsequent revision of his paper, but both men missed a presentation of A_6 as a subgroup of $\mathrm{SL}(3;\mathbb{C})$ subsequently found by Valentiner [1889] and named after him.

Finally, Klein gave a startling visual description of the Riemann surface. G_{168} cannot be the symmetry group of a surface in space, but the subgroups of order 24 mentioned above can be, for each is the symmetry group of an octahedron. The octahedral group permutes the four pairs of diametrically opposite points in the middle of the faces of the octahedron, and Klein showed that the octahedral subgroups in G_{168} permute 4 pairs such as A, A'; B, B'; C, C'; and D, D'. The 24 $\boldsymbol{a}$-points may be taken as the centers of 24 heptagons, and when the figure is cut up along the six heavy zig-zag lines the heptagons may be taken in 3's so

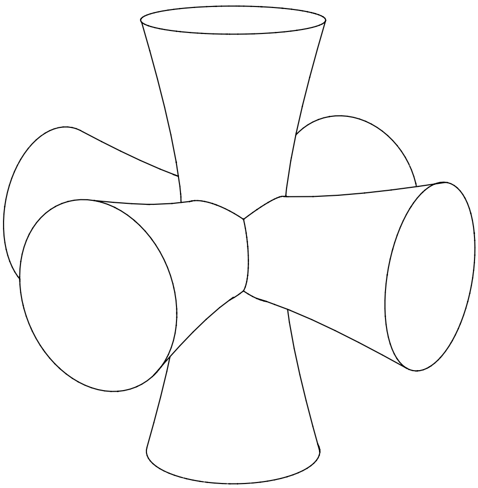


Figure 8

that each face of the octahedron is covered, more or less, with 21 shaded and 21 unshaded triangles. What is missing from this quilt is the six vertices of the octahedron. He regarded G_{168} as acting partly by rotating the octahedron, and partly by sliding the quilt over the octahedron, using identifications across the edges surrounding the vertices, which come from this dissection. He showed that these identifications were of diametrically opposite points, and were best performed by supposing the edges drawn out to infinity. He invoked the analogy with the hyperboloid of one sheet, which, projectively, is a torus. The dotted curve in Figure 1 represents the intersections of the curve with the plane at infinity.

So he described the surface as three hyperboloids whose axes meet at right angles, which is certainly appealing (see Figure 8).

A few remarks should be made about why Klein studied this problem in the first place. An old problem in the theory of elliptic functions asks for a relation between the moduli of elliptic integrals if the corresponding ratios of the periods is increased by a prime p. Jacobi and Abel had shown that the moduli were then related by a polynomial equation of degree $p+1$. Galois know that the polynomial equation could be reduced in degree to an equation of degree p when p was 5, 7, and 11, but for no higher prime, and he know that for these equations the corresponding group permuting the roots (its "Galois" group, as we say, for this reason) was $\mathrm{PSL}(2, \mathbb{Z}/p\mathbb{Z})$. Various mathematicians, notably Betti [1853] and Jordan [1868], attributed this reduction to the existence of large subgroups (of index p) in $\mathrm{PSL}(2, \mathbb{Z}/p\mathbb{Z})$ when $p = 5, 7, 11$, and showed that such subgroups did not exist for higher p. For example, when $p = 7$, the group permutes its 7 conjugate octahedral subgroups, so it has a permutation representation of degree 7.

Dedekind's paper [1878] was devoted to establishing a theory of modular functions without recourse to the existing theory of elliptic functions, and central to it was a function val : $H \to \mathbb{C}$ (from the German *Valenz*) which takes each value once and only once on the interior of R (and "half" its boundary) and for which

$$\operatorname{val}(z) = \operatorname{val}\left(\frac{az+b}{cz+d}\right).$$

Dedekind obtained it by taking the hypergeometric differential equation

$$x(1-x)\frac{d^2y}{dx^2} + (\gamma - (\alpha+\beta+1)x)\frac{dy}{dx} - \alpha\beta y = 0,$$

where x and y are complex, and so choosing α, β, and γ that a quotient of the two solutions maps $\mathbb{C}$ on to R or a copy of R under the action of $\mathrm{SL}(2;\mathbb{Z})$. Then this function val is the inverse of this quotient.

Klein took over Dedekind's val function and renamed it J. He also took over (with due acknowledgement) Dedekind's theory of modular transformations, in which $J(z)$ and

$$\tilde{J}(z) = J\left(\frac{Az+B}{Cz+D}\right)$$

are related, where $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ is a 2×2 matrix with integer entries and determinant p (or, more generally, any natural number). It is not hard to see that if we regard as equivalent the elements

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

where $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2;\mathbb{Z})$, then there are $p+1$ inequivalent $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ of determinant p. So the equation relating J and $\tilde{J}$ is a polynomial equation of degree $p+1$: the modular equation. Klein's contribution lay in making the groups explicit, and, more importantly, in introducing the Riemann surface of $\tilde{J}$ spread out over the complex J-sphere. Thus when $p = 5$ a sphere is obtained, corresponding to the role of $\mathrm{PSL}(2;\mathbb{Z}/5\mathbb{Z})$ as the icosahedral group, and we have seen detail how Klein treated the next case, $p = 7$. Indeed his famous book on the icosahedron [Klein 1884] is a showcase of his ideas on mathematics at the time, and his treatment of the quartic curve marks a high point in his style.

Rational functions in J are quotients of polynomials in J, and form a field $\mathbb{C}(J)$. Rational functions in $\tilde{J}$ live on the Riemann surface for $\tilde{J}$, and form a field extension of $C(J)$ whose Galois group is the group of the modular equation. This work of Klein's is thus at the origin of the Galois theory of function fields. Gordan's jocular name for this kind of mathematics was, Klein tells us, hyper-Galois theory [Klein 1922, p. 261]. The Galois-theoretic point of view is slightly further developed in Klein and Fricke's two-volume work on modular functions [Klein and Fricke 1890–92].

Dedekind had been the first in Germany to lecture on Galois theory, and the first to stress the importance of the concept of an abstract group (in lectures around 1858; see [Purkert 1976]), but he chose not to stress these ideas here, and it was left to Klein and his students to develop them.

Finally, why did Klein not notice the connection with non-Euclidean geometry? There can never be a simple answer to this question, but Klein's preferences were for projective geometry, and for using group theory to get at the invariant configurations (inflection points, bitangent points, and so on). In the paper analysed here, and in his work on the icosahedron, he succeeded brilliantly in his chosen task. The differential-geometric approach to non-Euclidean geometry advanced by Beltrami was less congenial to him, less central to his view of mathematics. So he did not look for such aspects of his problem, and the simple realization was left to Poincaré — to dramatic effect. Nonetheless, as Klein tells us [1923, p. 584], it was during a sleepless night, March 22–23, 1882, that Klein, in contemplating Figure 1, was able to grasp the full generality of Poincaré's ideas and so to formulate his own approach to automorphic functions and the uniformization of Riemann surfaces.

References

[Betti 1853] E. Betti, "Sopra l'abbassamento delle equazioni modulari delle funzioni ellittiche", *Annali di scienze mat. e fis.* (comp. da Barnaba Tortolini) **4** (1853), 81–100.

[Clebsch 1864] A. Clebsch, "Ueber die Anwendung der Abelschen Functionen in der Geometrie", *J. für reine angew. Math.* **63** (1864), 189–243.

[Coble 1908] A. Coble, "A configuration in finite geometry isomorphic with that of the 27 lines on a cubic surface", *Johns Hopkins University Circular* **7**:80-88 (1908), 736–744.

[Coble 1913] A. Coble, "An application of finite geometry to the characteristic theory of the odd and even theta functions", *Trans. Amer. Math. Soc.* **14** (1913), 241–276.

[Dedekind 1878] R. Dedekind, "Schreiben an Herrn Borchardt über die Theorie der elliptischen Modulfunctionen", *J. für reine angew. Math.* **83** (1878), 265–292. Reprinted as pp. 174–201 in his *Gesammelte mathematische Werke*, vol. 1, Vieweg, Braunschweig, 1930; reprinted by Chelsea, 1969.

[Dickson 1900] L. E. Dickson, *Linear Groups, with an exposition of the Galois Field Theory*, Chicago, Leipzig, 1900. Reprinted by Dover, New York, 1958.

[Geiser 1869] C. F. Geiser, "Ueber die Doppeltangenten einer ebenen Curve vierten Grades", *Math. Annalen* **1** (1869), 129–138.

[Haskell 1891] M. W. Haskell, "Ueber die zu der Curve $\lambda^3\mu + \mu^3\nu + \nu^3\lambda = 0$ im projectiven Sinne gehörende mehrfache Ueberdeckung der Ebene", *Amer. J. Math.* **13** (1891), 1–52.

[Hesse 1844] L. O. Hesse, "Ueber die Wendepunkte der Curven dritter Ordnung", *J. für reine angew. Math.* **28** (1844), 97–106. Reprinted as pp. 123–136 of his *Gesammelte Werke*, Chelsea, New York, 1972.

[Hesse 1855] L. O. Hesse, "Ueber die Doppeltangenten der Curven vierter Ordnung", *J. für reine angew. Math.* **49** (1855), 279–332. Reprinted as pp. 345–404 of his *Gesammelte Werke*, Chelsea, New York, 1972.

[Jacobi 1850] C. G. J. Jacobi, "Beweis des Satzes, daß eine Curve n^{ten} Grades in Allgemeinen $\frac{1}{2}n(n-2)(n^2-9)$ Doppeltangenten hat", *J. für reine angew. Math.* **40** (1850), 237–260. Reprinted as pp. 517–542 of his *Gesammelte Werke*, vol. 3, Gauthier-Villars, Paris, 1884.

[Jordan 1868] M. C. Jordan, "Note sur les équations modulaires", *C. Rendus Acad. Sci. Paris* **66** (1868), 308–312. Reprinted as pp. 159-163 of his *Œuvres*, vol. 1, Gauthier-Villars, Paris, 1961.

[Jordan 1870] M. C. Jordan, *Traité des substitutions et des équations algébriques*, Gauthier-Villars, Paris, 1870.

[Jordan 1878] M. C. Jordan, "Mémoire sur les équations différentielles linéaires à intégrale algébrique", *J. für reine angew. Math.* **84** (1878), 89–215. Reprinted as pp. 13–140 of his *Œuvres*, vol. 2, Gauthier-Villars, Paris, 1962.

[Klein 1879] F. Klein, "Ueber die Transformationen siebenter Ordnung der elliptischen Funktionen", *Math. Annalen* **14** (1879), 428–471. Reprinted as [Klein 1923, LXXXIV, pp. 90–136]. Translated in this collection.

[Klein 1882] F. Klein, *Ueber Riemanns Theorie der algebraischen Funktionen und ihrer Integrale*, Teubner, Leipzig, 1882. Reprinted as Reprinted as [Klein 1922, XCIX, pp. 499–573] Translated as *On Riemann's Theory of Algebraic Functions and their Integrals*, Macmillan and Bowes, Cambridge, 1893; reprinted by Dover, New York, 1963.

[Klein 1884] F. Klein, *Vorlesungen über das Ikosaeder und die Auflösung der Gleichungen vom fünften Grade*, Teubner, Leipzig, 1884. Translated as *Lectures on the icosahedron and the solution of equations of the fifth degree*, Kegan Paul, London, 1913 (2nd edition); reprinted by Dover, 1953.

[Klein 1922] F. Klein, *Gesammelte Mathematische Abhandlungen, 2: Anschauliche Geometrie; Substitutionsgruppen und Gleichungstheorie; zur mathematischen Physik*, edited by R. Fricke and H. Vermeil, Springer, Berlin, 1922. Reprinted by Springer, 1973.

[Klein 1923] F. Klein, *Gesammelte Mathematische Abhandlungen, 3: Elliptische Funktionen* etc., edited by R. Fricke et al., Springer, Berlin, 1923. Reprinted by Springer, 1973.

[Klein and Fricke 1890–92] F. Klein, *Vorlesungen über die Theorie der elliptischen Modulfunctionen,* ausgearbeitet und vervollständigt von Robert Fricke (2 vol.), Teubner, Leipzig, 1890–92. Reprinted by Johnson Reprint, New York, 1966.

[Plücker 1835] J. Plücker, *System der analytischen Geometrie, auf neue betrachtungsweisen gegrundet...*, Duncker und Humblot, Berlin, 1835.

[Plücker 1839] J. Plücker, *Theorie der algebraischen Curven: gegrundet auf eine neue Behandlungsweise der analytischen Geometrie*, Adolph Marcus, Berlin, 1839.

[Poncelet 1832] J. V. Poncelet, "Théorèmes et problèmes sur les lignes du troisième ordre", *Quetelet, Corr. Math.* **7** (1832), 79–84.

[Purkert 1976] W. Purkert, “Ein Manuskript Dedekinds über Galois-Theorie”, *Schriftenreihe für die Geschichte der Naturwissenschaft, Technik und Medezin* (Berlin) **13**:2 (1976), 1–16.

[Riemann 1857] B. Riemann, “Theorie der Abelschen Functionen”, *J. für reine angew. Math.* **54** (1857), 115–155. Reprinted as pp. 88–144 in his *Gesammelte mathematische Werke*, Teubner, 1862; reprinted by Dover, 1953, and by Springer, 1990.

[Riemann 1862] B. Riemann, “Zur Theorie der Abelschen Functionen”, pp. 487–504 in *Gesammelte mathematische Werke*, Teubner, Leipzig, 1862. Reprinted by Dover, 1953, and by Springer, 1990.

[Steiner 1848] J. Steiner, “Ueber allgemeine Eigenschaften der algebraischen Curven”, *Berichte der Akad. Wiss. zu Berlin* (1848), 310–316. Reprinted in *J. für reine angew. Math.* **47** (1854), 1–6, and in *Jakob Steiner's Gesammelte Werke*, G. Reimer, Berlin, 1882 (reprinted by Chelsea, Bronx, NY, 1972).

[Valentiner 1889] H. Valentiner, *De endelige transformations-gruppers theori*, Danish Academy publications, series V **6**, Copenhagen, 1889.

[Weber 1876] H. Weber, *Theorie der Abel'schen Functionen vom Geschlecht* 3, G. Reimer, Berlin, 1876.

[Weber 1896] H. Weber, *Lehrbuch der Algebra*, Vieweg, Braunschweig, 1896. Reprinted by Chelsea, New York, 1961.

Jeremy Gray
Faculty of Mathematics
Open University
Milton Keynes, MK7 6AA
United Kingdom
j.j.gray@open.ac.uk

The Eightfold Way
MSRI Publications
Volume **35**, 1998

Eightfold Way: The Sculpture

HELAMAN FERGUSON
WITH
CLAIRE FERGUSON

Abstract. This article covers some of my thinking while developing *The Eightfold Way* and some of the physical processes I used in creating it. The sequence of topics followed is:

My View
Ramanujan–Michelangelo
Geometry–Topology
Counting–Philosophy
Geometry Center–MSRI
Two Stones
Athena–Escher
Robot–Stewart platform

The pictures, the text, and the references can all be read independently of each other.

Wheeled wheels of processes and thoughts form a sort of multidimensional torus embedded in our time and space. In this paper we survey a few of these satellites and their orbits about the sculpture called *The Eightfold Way*. This amounts to making explicit part of the mathematical environment when I finished this sculpture. I intend no mysticism here, only some shared furnishings of our minds and hearts — shared cultivation of our neuron and capillary landscapes.

A typical Helaman sculpture has layers of titles, ranging from a colloquial expression such as “eightfold way” at the top to precise mathematical symbols and syntax such as $x^3y + y^3z + z^3x = 0$ deeper down. The last equation describes the algebraic surface of Klein that inspired this sculpture. For some reason, maybe because there was going to be a sculpture at MSRI, maybe not, at about the end of May 1992, a lot of email correspondence on the Klein surface began among Thurston, Asimov, Osserman, Brock, Gross, Sibley, Kuperberg, Bumby, Clemens, Hirbawi, Mess, Grayson, Adler, Elkies, Riera, to list a few. The rich mathematical folklore that exploded via the internet led eventually to the publication of this book.

Dr. John Slorp, President of the Minneapolis College of Art and Design, once observed: "*The Eightfold Way* is the perfect biomorphic form: it is sensuous and intelligent at the same time." But for most people educated in traditional schools, mathematics comes across as anything but sensuous. My sculptures attempt to bridge this gap.

My View

Mathematics is an art form, which need not remain invisible [Hill et al. 1989; Simmons 1991; Cole 1998]. Some Math evokes art, some Art evokes math.

Art is a social event which the artist recognizes and sets up. People frequently ask me how long it took to do a particular sculpture. I answer by recalling my age at the time I finished that piece. Few people are satisfied with this humorous answer, but it really did take that long. Perhaps people ask this question because they wonder how long it would take them. This article answers the question and reveals how I go about doing a mathematical sculpture.

Sculpture has to occupy physical time and space, and aesthetically I consider time just as important as space. My sculpture involves the mathematical content of a timeless discipline. As a first response to its timeless aspect, I work in stone which took hundreds of millions or thousands of millions of years to form. Secondly, I work with endless geometries like tori or surfaces without boundary with no obvious beginning or end.

I have a practical motive for working in rock and stone. Stone, for all its potential beauty and age is common and worthless. These days it has small military value (though this was not always so; compare, for example [Avery 1966; Homer n.d., Book 16, lines 757-780; Holinshed 1587]). Metal, on the other hand, is still essential in warfare. A few thousand years ago the Romans appropriated the bronze sculptures of Athens to make war, and only a few decades ago the Nazis confiscated the bronzes of Paris to support their war. I prefer stone over iron, steel, and bronze. Why not use iron? Today we are iron rich, there is iron everywhere. But metal is still vulnerable — wait until a war in space creates a greater appetite for all metals. A stone sculpture without military value may extend the life of my art as social event.

I don't think that my choice of stone bears on the philosophical question of whether mathematical objects actually exist in some Platonic universe. The process of getting aspects of mathematical objects into our physical universe dramatizes fundamental things about this universe. For example, I carve by subtraction from an quarried piece of some geological formation. Subtraction makes the piece smaller and smaller. The chips and dust I make are not small compared to atoms and wavelengths of visible light. We do not live in a purely mathematical continuum universe; certainly continued subdivision breaks down. Our universe on a smaller scale is undulant with particles and lumpy with waves.

Slide #21

Figure 1. Top view of the Eightfold Way and the hyperbolic disc, taken from an upper window in the MSRI building.

I haven't thought seriously about doing sculpture on this scale. For now I am sticking with stone.

Mathematical theorems occupy neither physical time nor space but share the characteristics of a communication tool. Mathematics as a conceptual language has its own aesthetic. The language of mathematics has three remarkable features: abstraction, condensation, and prediction.

In mathematics we consciously choose the level of abstraction. Consider for example the idea of the group of symmetries which underlies the *Eightfold Way*. This group of symmetries can be thought of at two levels of abstraction: algebraic or geometric.

For condensation consider that vast tables laboriously computed (for example, [Spenceley et al. 1952; Luke 1977]) have been replaced by a single equation or algorithm encoded in silicon. Kepler replaced Tycho Brahe's tables of planetary orbits with simple equations of ellipses. Newton reduced Kepler's equations to simple derivations from the inverse square law.

Prediction is possible if some correlation can be established between mathematical abstractions and a physical situation. Engineers study hundreds of models for every airplane, bridge, or boat before construction. Physical model construction is expensive and time consuming. Mathematics provides a kind of ghost realm, which coupled with computer graphics, makes modelling quick and inexpensive. As a sculptor I work with this cheap ghost real estate and find the mathematical ghost language very helpful in designing sculpture. It may help that the sculpture has mathematical content, although sometimes this content creates additional difficulties in the form of new problems to solve [Cox et al. 1994; Ferguson and Rockwood 1993; Ferguson et al. 1996].

The usual access to mathematical ghost material is through the imagination or the use of computers. What you see on the two-dimensional computer screen is a very different thing when you have had studio hand eye experience. It is like the difference between watching underwater films without scuba experience and then watching underwater films having had scuba experience. There is no comparision. Computer graphics does not replace studio experience.

There has been a philosophical prejudice in mathematics against the use of pictures to communicate mathematical ideas. Lagrange was not the first to brag that his book contained no deceptive figures or drawings. It has been pointed out, however, [Barwise and Etchemendy 1991] that more people have probably been deceived by specious linear arguments than by two-dimensional pictures. (See also [Barwise and Etchemendy 1996].) Some middle eastern religions were quite explicit in barring images. These anti graven image attitudes also colonized thinking in Puritan America and persist in some quarters even today. Curiously enough there is a historical record of a reversal in conservative religions in which certain properly clothed people forms may be acceptable but abstract visual form is considered unquestionably questionable. Robert Hughes in his treatise on art in America [Hughes 1997] argues these matters compellingly. Such attitudes have been part of our recent past, yet the visual image, suitably qualified, enjoys a rebirth in all the sciences, particularly mathematics.

In the previous century many drawings of functions and dramatic plaster models of mathematical forms were made; see [Jahnke and Emde 1945; Fischer 1986]. Today, computer images replace the dramatic plaster models. Bill Thurston has made the observation that computer graphics enables mathematicians, who typically are not trained to draw well, to draw computer pictures to communicate their ideas visually. Alfred Gray fills his book [1993] with images which took many person years to create before computers. The images of Gray's book come from parametric equations which have been under design by mathematicians for hundreds of years. Now they appear on the computer screen in a few keystrokes by anyone who can type. However, the old plaster models have a three-dimensional immediacy that transcends an image on a computer screen.

Mathematics is timeless, conjectural, and minimalist. How old is a theorem? It seems timeless because once thought and concluded, it appears to have always existed. Conjecture, one of the most creative acts in mathematics, can be stated as simply asking the right question. Conjecture grasps limbs of the complex tree of possible deductions, but runs deeper than that. Intuition becomes vital because there are assertions which are true but not deducible within the ambient system [Adamowicz and Zbierski 1997; Blum et al. 1998]. Mathematics is minimalist to the point of being invisible. Very few people get to see the theorems. Often a lot of hard work is involved in seeing or understanding a theorem. Mathematicians tend to communicate their most sublime creative acts to only a few of the mathematically trained. Part of the reason for this solipsism lies in the inherent character of the discipline. First, it is a discipline, hence the hard work. My algebraic topology professor, Tudor Ganea, used to say that "mathematics progresses by faith and hard work, the former augmented and the latter diminished by what others have done." Second, the mathematician strips away every nonessential idea. This makes her or him a minimalist of ideas. Furthermore, she or he may create a new language to assist in this reduction [MacLane 1971]. Who can speak such ex nihilo language?

I think of one of my sculptures as moving across time and space, an accretion of secondary aesthetics, anatomy, concepts, history, mathematics, philosophy, and process. Suppose someone digs up my sculpture in 1000 years, or 10,000, or even a million years. Can an interesting chapter of our nineteenth to twenty-first century mathematics be derived from it? We live in a golden age of mathematical creativity which does not necessarily have to continue. I prepare my sculpture to evoke deep thought in future 10^t time.

As I travel around the country lecturing and exhibiting my sculpture, I am approached by various mathematicians who shyly confess their artistic side. I enjoy encouraging them. When I began doing mathematical sculpture three decades ago, I had no one to talk to, no guide. Art was art and science was science and the two didn't converse [Snow 1959]. In graduate mathematics classes I knew better than to reveal that I took graduate sculpture classes and vice versa. On the rare occasions when such facts leaked out there was usually some display of hostility from one side or the other. A lot has changed in thirty years for the better, and I have probably helped change the old attitudes, but my work is just a beginning. More theorems in sculptural form would advance public appreciation and understanding of mathematics.

By writing this I hope to expose a path for others by providing a couple of forms of encouragement. First, a description of processes that a budding mathematical sculptor could in principle follow — a guide. Second, to make it seem easy — this is a lie, but one I have believed myself frequently and persistently. I excuse this mind game because I have found that most of what we believe isn't true and that verity doesn't stop us from acting in either creative or destructive ways and then justifying those ways by our beliefs. Most of what any person believes is not regarded as true or even helpful by some others. That, however, doesn't stop us from acting effectively and passionately on our beliefs and thereby accomplishing worthwhile and inspiring things. Mathematics is part of my belief system [Davis 1994].

Ramanujan–Michelangelo

Mathematicians have a highly developed, if solipsistic, aesthetic of their own, which they seldom share. They seem pretty shy or emotional about this [Cannon 1991; 1996]. However, they sometimes express this aesthetic with analogies outside their own field. G. N. Watson offers a particularly striking and mystifying example; he was an analyst after the school of G. H. Hardy, the English mathematician who had a remarkable and close relationship with Srinivasa Ramanujan [Newman 1956, vol. 1, pp. 366–376]. Watson's example appeals to me because in academic life, I was a computational number theorist [Ferguson and Forcade 1979; 1982; Ferguson et al. 1998], and cut mathematical milkteeth in [Whittaker and Watson 1927]. I will describe how Watson introduces a sculptural example, coincidentally close to my early artistic milkteeth [Avery 1966; Beck et al. 1994; Poeschke 1996].

Ramanujan loved to write down well poised specific cases of very general mathematical identities, choosing aesthetically rich examples. He seldom gave proofs of these identities and the way he came up with them seems mysterious to most. Watson spent a good part of his mathematical work proving Ramanujan's identities and confessed that the following integral series identity of Ramanujan thrilled him.

$$\int_{0\le x<\infty} e^{-3\pi x^2}\,\frac{\sinh \pi x}{\sinh 3\pi x}\,dx = \frac{1}{e^{2\pi/3}\sqrt{3}}\sum_{n\ge 0} e^{-2n(n+1)\pi}\prod_{0\le k\le n}\left(1+e^{-(2k+1)\pi}\right)^{-2}$$

Do you get a buzz from this? I certainly did; enough so that I had to verify this identity for myself. For starters, the left hand side has a continuous integral, the right had side has a discrete sum. Watson compares his thrill to the time he stepped into the Medici Chapel in Florence and saw the tomb of Giuliano surmounted with Night and Day and the orthogonally placed tomb of Lorenzo surmounted with Dusk and Dawn. There are identities there too: Night and Day — female and male figures, Dusk and Dawn — male and female figures. The angular male summands of the right hand side of Ramanujan's identity are in counterpoint with the rounded female integrand of the left hand side. The integrand drawn as a function over the domain $-\frac{1}{3} < x < \frac{1}{3}$ could well be a template for parts of Night and Dawn.

When Claire and I stepped into the Medici Chapel early one morning in Florence, I did not think of Ramanujan's identities, but I was definitely aroused and stirred in new ways. I reacted to the work of both masters whose seamless work looks deciptively easy with the exhilarating thought that I could do this. But while I could prove or duplicate or participate in either in a constructive way, coming up with either work originally constitutes another matter entirely. Both are priceless gifts to our civilization, but we ourselves choose to prize the priceless. This choice has to be based upon discipline and hard work.

The integral series identity is a coupling of two things, computationally representing very different algorithms, but actually the same. Both sides of this identity give the same real number, an infinitude of decimal digits beginning with

$$0.065329585951878667568200378714415958115553432840669156288403\,23\ldots$$

The duality, the coupling of dissimilar things which unite in some way is a strong formal theme in many works of art and science. The Night and Day, Dusk and Dawn, first female and male, then male and female nude figures is an explicit duality. Each represents different activities, sleeping and awake, dozing and arousing. These nudes provide a very human and very alive counterpoint to the dessicated corpses inches away under their forms. Yet the nudes are dead stone. Their aliveness centers wholly in the observer, just as Ramanujan's identity lives in an acutely sensitive reader, without whom his work remains dead ink on a page. Watson senses within himself a moving relationship between the creation of a poor Indian clerk from Kumbakonam near Madras and the creation of a semi-orphaned Italian stone cutter from Settignano near Florence. That Watson shared this duality confirms my aesthetic perception of mathematics and sculpture as relatible forms. Chandrasekhar, a mathematical physicist, reported on Watson's feelings in his book on truth and beauty [Chandrasekhar 1987], by devoting a part of a page to the integral-series identity of Ramanujan and two whole pages to the male-female images of Michelangelo.

In *The Eightfold Way* I used a number of formal dualities similar to those in the Ramanujan–Michelangelo relationship. To illustrate the pairing of two areas of mathematics, geometry and topology, I couple a black two-dimensional platform with a white three-dimensional tetrahedroid. In a later work I couple three-dimensional red and black Klein bottles even more explicitly so that they orbit each other over a platform of two-dimensional multiple images of themselves [Blankstein 1998; Cipra 1997]. [Senechal 1996] contains an illustration of another two-dimensional platform and correlated three-dimensional sculpture subsequent to *The Eightfold Way.*

Geometry–Topology

Mathematical elements of topology and geometry, united by group theory comprise environmental influences upon *The Eightfold Way.* The white Carrara marble tetrahedroid is a topological statement. This is not, strictly speaking, a tetrahedron. I carved it in a qualitative free form process known as direct carving, paying attention to the combinatorics and topology but not rigid or measured geometry. By contrast, the black and green serpentine hyperbolic disc tiled platform base is a geometric statement. This is quantitative, everything is measured carefully to preserve the rigid geometry. Indeed, I carved this part using a computer driven water-jet robot, driven by a straight line program foloowing coordinates of explicit numbers. The black serpentine prism creates a connecting homotopy from the regular 120-degree hyperbolic geometric heptagon in the base platform to the topological palm of a hand shaped heptagon supporting the tetrahedral form. This junction prism provides the transition from quantitative accuracy (one-thousandth of an inch in this case) to qualitative flowing elusive forms.

The expressive relationship between the two, the topology and the geometry was very important to me. I was pleased when I installed the heptagonal prism or pedestal upright in the center of the hyperbolic disc and Bill Thurston came out and saw it for the first time, remarking, “that really is Topology”.

By the expression *quantitative* in a sculpture, I mean that overall accurate measurements are important: exact angles, exact lengths. I regard the expression of a distance function and the specific point to point length relationships as essential to the reading of the form. Rigid relationships are important, but not necessarily the physical scale, in this context [Apéry n.d.; Sequin n.d.]. Once an object is made we can measure it and it becomes quantitative. But here I am speaking about the original process leading to the design and creation of the final object.

By the expression *qualitative* in a sculpture, I mean that specific measurements were not an important part of the process of creating it. The expression of the topological or combinatorial features, however, are paramount. The form idea is invariant under smooth deformations. I make quantitative choices in smooth deformations for reasons other than the mathematical reading of the form. An observer of sculpture tends to want to touch qualitative but not quantitative. Marble provides a medium for qualitative expression, as it polishes to a reflective sheen which pleases the eye and pleasures the touch. Running a hand over the grooves and surfaces of *The Eightfold Way* provides an unforgettable sensual experience.

One general theme reflected in this sculpture is that rigid geometry has an underlying topology and vice versa: a way to look at topology is to look for an underlying rigid geometry. This theme arose repeatedly in the last couple of centuries in mathematics, from Euler to Poincaré, Klein, and Fricke. It recently has been developed extensively in the work of William Thurston and others; see [Thurston 1997; Ratcliffe 1994], for example.

The Eightfold Way directly addresses the symmetry of surfaces: A sphere with $g > 1$ handles cannot have infinite symmetry, whereas a sphere $g = 0$ and a torus $g = 1$ both have infinitely many symmetry preserving transformations. Hurwitz' Theorem [1893; 1987] gives an upper bound on the symmetry: the group of automorphisms of a surface of genus $g > 1$ is bounded by $84(g - 1)$. The surface offered by the marble has genus $g = 3$ which in this case is a tetrahedral form with four faces, each face penetrated by the ambient space so that all penetrations meet in the middle.

This tetrahedral configuration appears to have four "handles" corresponding to the four edges of the tetrahedron. That these four are really three handles could confuse non-mathematicians. Claire's immediate response was, "Oh, that's easy, it's just like having a baby, you make a great big open smile here and then you see this head with two eyes and a mouth". (Claire knows the topology of having babies by heart, inside out and backwards.)

To make sense out of Claire's response, take a piece of soft clay and deform it into a tetrahedron, press holes into the four sides so they all meet in the middle, then without tearing deform open one of the triangular holes and flatten everything until three holes are visible. Another way to see this is to make three cuts through three limbs until the form becomes a ball with knobs and no loops.

The automorphism or symmetry group of a surface of genus three can have as many as $84 \cdot (3 - 1) = 84 \cdot 2 = 168 = 24 \cdot 7$ elements. I hear a lot of talk in the software world about deadlines being met by a $24 \cdot 7$ effort of twenty-four hours seven days a week, a never ending symmetry in time. Everyone experiences the stretched out symmetries of 24 hours in a day, 7 days in a week, and 168 hours in a week. The choice of the prime factors 2, 3, 7 in our organization of time keeping is ancient and interesting in its origins of oversimplifying of natural events [Neugebauer 1975]. The 2 is structural, diurnal, day and night, but why the 3 and 7?

What does such symmetry mean in a spatial physical or sculptural context? The mathematical definition of symmetry cannot be taken literally because of fundamental physics usually expressed by Heisenberg's uncertainty principle. It is impossible to manufacture a large object with precisely matching parts or exact symmetry. We come mechanically, molecularly, or even visually close; we come dramatically close in the case of cutting diamonds. We don't touch the symmetry of diamonds, we keep them small and wear them instead. If they were measured in pounds and feet instead of carats and millimeters their ever present imperfections would be grossly evident.

Symmetry in our world approximates perfection at best and a deception at worst. In the case of *The Eightfold Way*, the literal group of symmetries of the polished marble surface itself has only one element. The group of symmetries is trivial which means no symmetry at all. By doing sculpture in physical materials invariably all the symmetry gets broken, so symmetry has to be implied. Broken symmetry implies that a symmetry in theory is there to be broken [Morrison 1988]. By implication, *The Eightfold Way* surface articulates all 168 automorphism elements. More handily and literally, the generators can be read out of it.

Two- and threefold symmetries are implied by the tetrahedral form. The heptagon covering implies sevenfold symmetry. Each heptagon vertex forms a triple point or triskelion with the edges of its three neighbor points. The grooves or ridges of the three edges are curved to meet the neighbor point. There are 56 points and 84 edges to make up 56 triskelions in all. In carving this marble, I used a small plexiglass equilateral triangle form as a pattern to keep these triskelions under some equianglular control. This was a loose qualitative 120-degree consideration which rhymes with the more exact quantitative 120-degree triple points of the base platform.

This same symmetry reads more literally in the quantitative two-dimensional base platform. There the triple points are embedded in a system of infinitely many triple points. An infinite discrete group associates with this platform. This infinite group acts by hyperbolic transformations on the hyperbolic plane and has a fundamental domain of exactly 24 heptagons. In this case, there are 23 darker heptagons grouped around the 1 dark polished stone heptagonal prism in the center of the hyperbolic disk. The triskelions have been cut to have metrically accurate angles of 120 degrees. The discrete group transforms the fundamental domain in such a way that certain edges are identified. The transformations sew up the 24 heptagon domain into a surface of genus three, viz., into the marble surface lying above the fundamental domain. The boundaries of the 24 white marble heptagons carved into the tetrahedral form are articulated as either ridges or incisions. The incisions or cuts define the doubled outside boundary of the lower fundamental domain. The ridges on the marble also (compared to the incisions just described) form edges of heptagons. These ridges correspond to the geodesic arcs in the hyperbolic plane which lie inside the fundamental domain or cluster of darker contiguous heptagons. The day I installed this piece, Bill Thurston came out and started pasting notated tape on the white marble heptagons and connecting them with string to the corresponding black serpentine heptagons. The sun got too hot or something interrrupted this project. Some of the photographs taken at the time show these tapes.

If the viewer "reads" the sculpture a certain way the reason for the title "Eightfold Way" becomes quite clear. To "read", select an edge somewhere on

Slide #20

Plate 1. *The Eightfold Way* seen from the MSRI library (facing East).

Red/green/white circle

Plate 2. Thurston's rendition of the heptagon tiling.

Slide #6

Plate 3. Water-jet robot cutting out one of the serpentine block for the hyperbolic tiling. This set-up, part-holding, and cutting process had to be done over 232 times.

Slide #8

Plate 4. Full set of stone tiles as cut by the water-jet robot. Note the dark cluster in the middle and the corona stones on the rim. The cluster "sews up" into the tesselated tetrahedroid.

Slide #17

Plate 5. The artist doing finishing work during the installation in August 1993. Bill Thurston and Joe Christy stand behind.

Slide #25

Plate 6. Cathedral view from inside *The Eightfold Way*.

the white marble tetrahedral form. Go along this edge to the fork in the road and take the left fork. Go to the next and take the right fork, then the left fork, then the right fork, left fork, right fork, left fork, right fork. If the viewer counted carefully, she is back on the starting edge. There were eight turns at eight forks in the road, hence the title.

The left right path is a cycle because it returns. Cycles like this are called Petrie nets. In general, a Petrie net in some fixed polyhedron is a skew polygon where every two but no three consecutive sides belong to the same face of the polyhedron. Petrie nets were named by H. S. MacDonald Coxeter after John Flinders Petrie [Coxeter 1973], the only son of the great founding archaeologist Sir William Matthews Flinders Petrie, who studied pyramids in Egypt [Drower 1985]. These Petrie cycles correspond to powers of products of generators (commutators) of the 168 element group of automorphisms. There 21 such cycles possible among the 56 triple points, since each path returns after eight alternating turns to the inital choice. In reading this sculpture, we visualize the full symmetry genus three surface with more than our eyes, we have our fingers touching the stone along these eightfold paths. The human haptic sense of around and through becomes a vital supplement to seeing, perceiving and certainly enjoying a symmetry from more dimensions than we usually experience.

By coincidence, the group of 3×3 invertible and commutator matrices with entries over the two element field F_2 consisting of $\{0, 1\}$ has exactly 168 elements, $(2^3 - 1)(2^3 - 2)(2^3 - 2^2) = 7 \cdot 6 \cdot 4 = 7 \cdot 24 = 168$. This set of elements is a group denoted by $\mathrm{GL}(3, 2)$. There are 21 elements of order two and 56 elements of order three in this group. Does this correspond to the 21 Petrie cycles and the 56 vertices?

There are three abelian groups of order 168, and two nonabelian groups of order 168 of which only one is a simple group, viz., this $\mathrm{GL}(3, 2)$ is given by generators as $\langle a, b \mid a^2 = b^3 = [a, b]^4 = (ab)^7 = 1\rangle$, where $[a, b]$ is defined to be the commutator $[a, b] = aba^{-1}b^{-1}$. The relation $[a, b]^4 = 1$ is precisely the origin of the eightfold way path. Specific generators that satisfy these relations for $\mathrm{GL}(3, 2)$ are

$$a = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

This group $\mathrm{GL}(3, 2)$ has the polynomial

$$\eta(t) = 1 + 21t + 56t^2 + 42t^4 + 48t^7,$$

where the coefficient of t^n is the number of elements in the group of order n; moreover $\eta(1) = 168$, $\eta(-1) = -2 \cdot 21$.

Counting–Philosophy

This sculpture involves counting. New environments sometimes make counting difficult. It is easy to lose track along a given eightfold path so concentration is required. Counting was a major cultural and scientific achievement. Our systems of counting are very old — we have no real idea just how old. “No class of words, not even those denoting family relationship, has been so persistent as the numerals in retaining the inherited words” [Buck 1949, Chapter 13; Pappas 1994, pp. 145, 191]. Once one has learned to count, it seems the integers were there all along. Are they infinitely old?

As I carve these counting opportunities in stone, I wonder if maybe the stone is older than our counting. In fact, the age of the stone I carve is some fraction of a billion years. My process of direct carving seems to be moving back in time, as I reveal layer after layer of stone, deeper into our earth or solar system past, time and space coagulate and congeal. Mathematics seems timeless, and stone seems timeless. This is one reason I think stone is a natural material to express mathematical ideas.

Cultures that count cover the surface of the earth. World languages have been organized by linguists along continental and island structures where people have settled, isolated, and preserved their old counting ways. I believe it is possible to count orally to eight in exactly twenty one language groups, a different language group for each Petrie cycle and at the same time pretty much cover our globe. There seem to be 84 fairly different languages available for this purpose. On the other hand our world collection of visual number symbols does not seem to be as rich as our phonetic symbols. However, there is a set of Mayan heads that represent the first eight counting numbers. These were carved in stone so it seems I am not alone in carving numbers in stone.

The expression “Eightfold Way” has been recycled from similar titles. This may be appropriate given a sermon by the Buddha to potential disciples in a deer park near Benares or Varanasi [Kitagawa 1971; Levenson 1996; Powell 1995]. He summarized the practice of the Eightfold Path as the Three Learnings. This invites a confluence with my sculpture. The edges of each eightfold path on the sculpture are joined at each stage by a vertex of valence three. Each triple point in three dimensions above and in two dimensions below can be thought of as the Three Learnings organizing the path: moral precepts (sīla) encompassing speech, action, livelihood; meditative practive (dhyaña) encompassing effort, mindfulness, concentration; initially faith and ultimately attainment of wisdom (prajñā) encompassing understanding (view) and thought.

The triple points, and maybe old style triskelions, themselves are anatomical especially among peoples who recycle the bones of their dead; for example, skull offering bowls [Nomachi 1997, p. 147], kapala, from life in arid, rocky mountainous plateau regions (to dig is hard and unprofitable) include various anatomical references. Consider the sagittal and coronal sutures, anterior and posterior fontanelles. See [Levenson 1996, pp. 54-55] for eightfold path pictures as well as the triskelion images of buddhist skull bowls showing the common anatomical triple points in the skull, anterior: sagittal suture and two coronal sutures, triple junction; posterior: sagittal and two lambdoidal sutures. These type triple points also occur in the wheel of the law images [Netter 1996; Gray 1901; Richer 1890]. The human skull, cleaned, is a natural visual source of the triskelion form [Chumbawamba 1997].

The Eightfold Path was a paradigm designed for novitiates, a learners' course for beginners. At a certain stage of development there are ten: right concentration leads to perfect insight and perfect deliverance (two more), the end of the noble path. A sculptural analogy would be that at the end of eight edges is a triple point: after right concentration or meditation, one is faced with two new edges bordering on an entire heptagon. This leaves the linear or one-dimensional path to a field of two dimensions or infinitely many paths and perhaps insight and deliverance.

Sand paintings of the Eightfold Path involve arcs of circles in a disc arrangement suggesting a connection with the hyperbolic disc [Levenson 1996, p. 22]. Indeed there are mirrors that I have seen at the Art Institute of Chicago where the back of the mirrors have hyperbolic-like arrangements of arcs of circles. They are made with Dragon arabesques, Eastern Zhou Dynasty, Warring States period or early Western Han Dynasty, 3rd/2nd century B.C. Could these have Eightfold Path origins?

Religious or theological systems of thought tend to be very abstract systems of thought. Perhaps they represent some of our earliest mathematical forms. Much of history, especially the history of conflict reflects systems of thought. Even though they are based on very different, in some cases mutually inconsistent axioms, they have provided ample opportunity to go to war. On the other hand, I myself commit considerable violence in the process of carving stone. My abstract systems of mathematical thinking clash with the geology and mineralogy of the stone as I reform it in my own images. I war on my stone with hammers, chisels, diamond saws, grinders, and I have made the time honored excuses and impose my abstract thoughts with compelling violence.

The phrase "The Eightfold Way" is also the title of a book by Gell-Mann and Ne'eman [Gell-Mann and Ne'eman 1964], referring to an earlier paper by Gell-Mann where quarks where introduced theoretically by assuming three base states considered to transform according to the eight-dimensional group SU(3). (See [Lichtenberg 1978, pp. 166–171; Schensted 1976, pp. 218-228] for an exposition.) One of the systems of weight points for an irreducible representation of SU(3) corresponds to an octet of baryons including the proton and neutron. The dynamics of elementary particle interaction (scattering) is not well understood, so even approximate symmetries are vital to making predictions. Unitary transformations are associated with conservation laws, and the matrix group SU(3) provides approximate symmetries. The eight comes from SU(3) and the representation dimensions beginning with $1, 8, 10, 27$, where the 8 cooresponds to the most frequent higher mass of the vector baryons or mesons. Just to have some idea of how approximate this symmetry is from a mass perspective, consider the variation in the masses of the baryon octet, $(p, n, \Lambda, \Sigma^+, \Sigma^-, \Sigma^0, \Xi^0, \Xi^-)$ with masses $(938, 940, 1116, 1189, 1192, 1314, 1321)$. Other than the presence of the two 3×3 matrix groups $GL(3, 2)$ and $SU(3)$ any relationship between the physics eightfold way and the sculpture eightfold way remains unexplored.

Geometry Center–MSRI

A key fact behind the existence of any larger sculpture is funding. Elwyn Berlekamp had facilitated some funding for an unspecified MSRI sculpture from the Mitsubishi Electric Research Laboratories in Cambridge, Massachusetts. This vaguely had something to do with Kaplansky's retirement. Kaplansky was then Director of MSRI. The sculpture was originally going to be a development of the circle of theorems around the $(2, 3, 7)$ pretzel knot. This knot had begun its mathematical life with Seifert [1934, Satz 6, p. 589] computing its Alexander polynomial, then later came number theory connections with Lehmer [1933] and much since [Reid 1991; Riley 1975]. My granite $(2, 3, 7)$ pretzel knot sculpture has yet to see the light of day. Kaplansky's mathematical work has resisted sculptural expression so far.

I have a note from September 1990 about a chat with Bill Thurston, who said he would bring up the idea of one of my sculptures at MSRI. Then at the AMS–MAA meeting in Baltimore in January 1992 I talked with Lenore Blum, Deputy Director at MSRI, about the suggestion and she like the idea. The 18-month gap between these two events is typical of developing my sculpture. One has to be patient.

In March 1992 I FedExed a video and some posters to Bob Osserman, also Deputy Director at MSRI. We discussed some of Kaplansky's work and also the Lehmer conjecture. Later in the month Arlene Baxter, manager at MSRI, designer of the MSRI brochure, sent some photographs of possible sites for a sculpture. Bill was gone when I visited MSRI a few weeks later, but I sketched a brief idea on Bob's blackboard. This was a concept based on my findings at the Geometry Center in Minnesota during the great Halloween blizzard of a year or so before.

After Al Marden, director of the Geometry Center, saw *Knotted Wye I* at my exhibition at Ohio State University, in August 1990, he said they really needed a sculpture like it at the Geometry Center at the University of Minnesota. He felt that many people at the University there did not understand mathematicians. He had observed that people thought they were just computer hackers because the Geometry Center used heavy computer graphics as a research tool for gaining insight into geometry. He thought if they had a sculpture like *Knotted Wye I* that people would understand through it that they were mathematicians — artists of a certain kind. Creative mathematicians tend to think of their science as an art form, perhaps the ultimate conceptual art form (even if canonized academically in some ways).

How did I get involved with Bill Thurston and the topology and geometry themes of *The Eightfold Way*? In 1991 Don Davis at Lehigh University invited me to give a math sculpture talk. This was followed by another talk at the Five Colleges Geometry seminar at the University of Massachusetts at Amherst sponsored by Donal O'Shea and Lester Senechal of Mount Holyoke College. When Claire and I give such talks, I usually haul along some examples of smaller sculptures. This time I included the first Knotted Wye. (This knotted wye hyperbolic theme had been mentioned to me by Bill Thurston at the AMS-MAA meeting in Boulder, he did a clay sketch which Gary Lawlor, a post-doc at Princeton, brought down to me while visiting us in Maryland.) On the way back we stopped in Princeton and I showed Bill the bronze *Figureight Knot Complements, a Wild Sphere*, as well as the Carrara marble *Knotted Wye* (it didn't have a number then; cf. [Ferguson 1994]). At that time Bill mentioned a $\mathrm{PSL}(2,7)$ symmetry group surface problem and suggested that perhaps there was a sculpture there. This was the first hint of what eventually developed into *The Eightfold Way.*

The primes 2, 3, 7 of the pretzel knot reappear as the only prime factors of the order of the group $\mathrm{PSL}(2,7)$. The connection of $\mathrm{PSL}(2,7)$ with the Klein surface became more interesting to me sculpturally on the occasion of the dedication of one of my *Knotted Wye II* [Ferguson 1994]. This 1500-pound Carrara marble was preceded by the smaller *Knotted Wye I* mentioned above. Both are direct carvings. Their configuration can be decoded from a verbal description of the planar knotted graph presentation. The first link goes over, under, over, under, the second link under, over, and the third link over, under, over, under, going in each sequence from the first vertex to the second vertex. This knotted graph fits into the family of Kinoshita–Wolcott knotted graphs of k, m, n full twists; see [Farmer and Stanford 1996]. It appears in open ended but equivalent form in Ashley's Book of Knots [1944] as the wall knot and the further development of Matthew Walker's knot. There is an associated yarn of how the knot saved this sailor Walker from certain hanging. The judge was a former sailor and said he would let Walker off if Walker could tie a knot the judge had never seen. Matt tied his knot like a small fist in the middle of six fathoms of rope. The judge was impressed enough to give the sailor his freedom.

The dedication for *Knotted Wye II* illustrates the desire people have to experience mathematics in a direct way. As soon as the dedication was over there was a collective breath and the audience rushed forward to touch the marble carving. They climbed all around it, and held hands through the sculpture's limbs. These were adults, their spontaneity, lack of selfconsciousness, and involvement made it a delicious moment for me.

Knotted Wye II will communicate mathematics for many generations, sitting as it does on four oak cuboids. This is another instance of rigid geometry underlying the fluid topology, a precursor of the "Eightfold Way" concept described above. This work was first installed in the Geometry Center but has since been moved to the Mathematics Library at the University of Minnesota. It is the first 1500-pound theorem in the Frederick Weisman Museum of Fine Art Collection.

One of the interesting discussions that week at the Geometry Center was about the Klein surface and $\mathrm{PSL}(2,7)$. John Horton Conway showed me an amazing $\mathrm{PSL}(2,7)$ contiguity relationship for the Klein surface. He grabbed a scrap of paper and scribbled down the $\mathrm{PSL}(2,7)$ relations, group elements, the appropriate conjugacy classes and what I called the eightfold way relationship. It took me several years to convert this two-dimensional scribble, its implications, and some of its mathematical context into the eightfold way sculpture.

For most of this week Margaret Thurston was sewing up a patchwork of regular heptagons into John Conway's incidence scheme. Margaret's stuffed heptagons were still at the Geometry Center when we visited in the Fall. We were scheduled to give a slide lecture to a large group of high school math teachers and math students from all over Minnesota (The Humpty-Up program of Harvey Keynes). The Great Halloween blizzard of 1991 closed the airport and marooned us in a hotel, but I managed to wade to the Center through four feet of drifting snow. There I found Margaret's extraordinary stuffed object and started thinking about it. I found the heptagons were somehow wrapped around a tetrahedral skeleton which suggested two- and threefold symmetry. I made a foam version and carved some figureight knot complements in styrofoam. I left knot complements there, but brought the tetrahedral foam with its tesselation of heptagons back.

In May 1992 Bob Osserman showed my blackboard drawing to Bill Thurston who talked to John Horton Conway about $\mathrm{PSL}(2,7)$ as an MSRI sculpture. Bob came out in visited us in early May 1992. By this time I had a full scale size tetrahedroid carved out of white styrofoam with incisions to indicate the tesselation. By May eighth I settled on a hyperbolic prism to support the tetrahedroid. I wanted this piece to be approachable by an average size person and easily touched. Meanwhile, in our communications, Bob was pressing for a circular area maybe filled with sand in which to stand the sculpture. I had no plans to include a hyperbolic disc, this was where the Gauss problem came in.

The problem was Gauss, Gauss was the name of MSRI's resident cat. If the circle were filled with sand or raked white gravel Gauss might chose to appropriate the site as his personal cat box. This would discourage people from stepping close to the piece. I wanted to encourage people getting close enough to reach in and around the sculpture to follow the tesellation ridges and grooves. What to do? Gauss the Cat had to be respected. Gauss the Cat's namesake, Carl Frederick Gauss, had actually invented hyperbolic geometry perhaps even before Bolyai or Lobachevsky. It eventually occurred to us that if the Poincaré disc model in stone replaced the proposed sand or gravel then the Gauss problem would be solved. Gauss the Cat showed no proprietary interest in the Poincaré disc model of hyperbolic plane geometry and there would be no problem with

Slide #1

Figure 2. John Horton Conway's page of notes describing his $\mathrm{PSL}(2, 7)$ action on the Klein surface.

Slide #2

Figure 3. Conway's sketch of heptagon contiguity associated with the $\mathrm{PSL}(2,7)$ action. The heptagon ∞B became the joint between the white marble and black serpentine.

Slide #3

Figure 4. Michael Ferguson (our youngest son) wearing Margaret Thurston's stuffed version of the Eightfold Way tetrahedron. She had made a multicolored version in 1991, which Helaman studied during the Halloween Blizzard in Minneapolis.

people standing on the hyperbolic tiling. The hyperbolic platform required some pretty extensive logistics. The heptagonal prism of 120-degree angles had to fit the real size of the conformal Poincaré disc that would mathematically scale with the rigid central hyperbolic regular heptagon. The next difficulty was cutting the hundreds of heptagonal tiles in stone to a few thousands of an inch precision; I solved that problem by cutting with a water-jet.

Serpentine–Marble

The black stone in the hyperbolic platform base of *The Eightfold Way* is serpentine, a magnesium silicate mineral related to granite, a compacted mineral talc with a very small rhomboid crystal size. Also called steatite, serpentine comes in a wide spectrum of quality and hardness. The softer steatites, which may contain asbestos type fibers, are called soapstones. Some of the oldest artifacts known were carved from soapstone. Because this mineral is impervious to heat and chemicals, it is used to line steel furnaces, build efficient wood and coal stoves, and make laboratory table tops. Some varieties are as hard as granite but with a finer grain. I wanted one of these hard types a vein of which occurs in the Blue Ridge mountain area of Albemarle County in Virginia. Early May Claire and I brought a thirty four hundred block from a stone yard there. I cut the fluted heptagonal prism out of this block. The age of this serpentine has been estimated to be between 400 and 500 million years old.

The white stone for the tetrahedroid from posed an interesting size problem. I needed enough stone to rough carve a tetrahedron $2\sqrt{2}$ feet on a side. Did this need to be a block of white marble $2\sqrt{2}$ feet thick? I could not find among my stone suppliers any cube that thick. Fortunately a two foot thick block would suffice! A very convenient feature of tetrahedrons is that they are not as thick as they seem from the edge length. I was first impressed by this listening to a talk in 1966 by Tracy Hall, the first person to synthesize diamonds in the laboratory [Hall 1986; Nassau 1980]. His technique, standard production process now, was to use a tetrahedral press with high pressure rams focussed on a regular tetrahedron. In his talk he showed how he tightly pinched a cylinder or straw, each successive pinch orthogonal to the next, giving a string of tetrahedrons. A seemingly too big tetrahedron slides through a seemingly too small straw. Baby heads are sort of tetrahedral and they get through an impressively small birth canal, in similar fashion. A tetrahedron of edge length $2\sqrt{2}$ can be carved out of a $2 \times 2 \times 2$ cube of marble, so I really only needed a block two feet thick. After checking on availability up and down the east coast I did find a suitably thick block of white marble from importer Harold Vogel of Manassas, Virginia.[1] I went down, split out my $2 \times 2 \times 2$ foot cube and brought that back in my

[1] It is an Italian Carrara marble. I incorrectly described in [Ferguson 1994] as Imperial Danby Vermont marble because of its similarity in carving to the latter stone.

Slide #4

Figure 5. Silvio Levy used Mathematica to make a 6-page collage of the heptagonal tiling of the Poincaré disc, containing about 600 tiles. Much annotated by the artist, this drawing served as a model for the stone cutting. About half the tiles — all but the outer layer — made it into the sculpture; they are represented by 18 classes, each of a slightly different Euclidean shape. (See Plate 3 for the cutting of the tiles.)

4×4 truck with augmented undercarriage. The age of this marble from Italy is around 200 million years.

In October 1992, I visited MSRI and worked with Silvio Levy on finalizing the tile data for input to the completely different PC system which controlled the water-jet I was planning to use to cut the hyperbolic tiles. I had worked out my own Mathematica programs for tiling the Poincaré disc with regular 120-degree heptagons. Silvio had been through something like this before when he generated an automatic version of the type of Escher's *Circle Limit III* [Levy 1994; Escher 1989, p. 43; Escher 1982, pp. 97, 320]. He quickly adapted the Geometry Center word generation programs to extract the Postscript form data I needed for the water-jet programs. After the conference I went to supervise the water-jet cutting.

Architect Bill Blass produced the final concrete patio drawings in July 1992. Since this sculpture was being installed over the Hayward fault zone I worked out the structural issues with engineer Nellie Ingraham. We agreed finally on five internal solid steel rods. This required an extensive system of holes to be drilled in the fluted heptagonal serpentine prism and the matching marble. Physically matching the heptagonal hand of the white marble to the corresponding heptagonal hand of the black serpentine was a problem that had to be solved before these holes could be drilled.

Slide #7

Figure 6. To debug the water-jet program a full set of wooden blocks was cut and then assembled.

Slide #9

Figure 7. Hyperbolic "circle limit" of heptagons tiling installed and curing at MSRI. The earthquake stabilization hole in the middle is not yet drilled into the concrete pad and footings.

Athena–Escher

The base platform of *The Eightfold Way* makes a direct visual connection with the circle limit woodcuts of Escher [Escher 1989; 1995; 1992], as mentioned above. Remarkably enough, Escher had solved the problem of having limit tilings converge to a boundary triangle or a boundary square, but was stopped at a circle boundary. His dilemma was solved when he discovered hyperbolic geometry by making the acquaintance of H. S. MacDonald Coxeter and his work [Coxeter 1957; 1979]. Escher's wonderful circle limit woodcuts came after that and an unthinkable amount of painstaking labor. See also [Coxeter 1998; Emmer 1980].

Developing the circle limit of heptagons of *The Eightfold Way* was very different from Escher's technique of creating woodblocks for his prints. We have computer technology today that Escher would I believe have been delighted to use. I did use a computer directed water-jet to cut wood blocks of the heptagonal disc tiling and I did make canvas prints of the hyperbolic tesselation. It is certainly possible to take rubbings suitable for framing off the stone hyperbolic platform of *The Eightfold Way*. I personally encourage people to do those rubbings, they are easy to do. The stone tiling itself was so difficult to create that visitors lifting off versions of it to take home will share the joy of the thing. Escher drilled holes in his circle limit wood blocks to prevent more impressions from being made. There are no limits to the number of impressions to be taken from the circle limit of heptagons of *The Eightfold Way*.

It was important to specify exactly in computer form the hyperbolic tiling blocks. Since the 231 precision water-jet cut stone blocks or tiles were to form a tesselation of the Poincaré conformal disk model of the hyperbolic plane, Mathematica and other programs were written to develop inputs for the controller computer of the water-jet. A robot, the water-jet system responds only to a meticulously prepared set of instructions. To accomodate the circle at infinity or boundary of the disc there are 14 disc rim or corona stones. Interior to that are 217 hyperbolic heptagon stone blocks. Each heptagon has seven interior angles, each of 120 degrees. The 217 tile ensemble has 23 dark and 194 light serpentine heptagons surrounding a center prism with an exact conformal heptagon base. The tiles were set in May 1993 by Lajos Biczo. (Given the role of Janos Bolyai in the early history of non-Euclidean geometry [Greenberg 1993], it was appropriate to have someone of Hungarian heritage set this non-Euclidean geometry disc.) Joe Christy of MSRI facilitated the setting of the tiles and protected them while the grout cured. The rest of the sculpture could not be installed until the curing process was complete.

One of the tiles, the center tile, is actually a prism. It relates not to Escher but to Athena. καλὸς κ'ἀγαθός (read "kalós kagathós"), the beautiful and the good, a Greek saying applied to people whose outer beauty reflected internal moral goodness. I want my sculpture to outwardly reflect the internal integrity and consistency of mathematical theorems. A reflexion of this theme appears more

Slide #5

Figure 8. Styrofoam maquette showing ∞B heptagon matching by SP-2, tetrahedron ∞B input, prism base ∞B output. This was to check the digitizer vs. inverse digitizer software of the SP-2 written in C by Sam Ferguson.

Slide #10

Figure 9. The 200 million year old white marble cubical block at a stage of being carved into its tetrahedral form.

Slide #13

Figure 10. SP-2 fitted maquette. A cloud of points for the surface of ∞B in the previously carved stryofoam tetrahedroid was rotated and became a virtual image target to create the other side of ∞B in the stryofoam base. This was to check the procedure before carving the stone version.

literally, in that I have included in the heptagonal prism a three-dimensional quotation from a fifth century Greek work which occurs in a series of twelve high relief metope over entrances to the temple of Zeus at Olympia. These sculptures, dated about 460 B.C., feature the labors of Herakles, the legendary founder of the Olympic Games, known to the Romans as Hercules. The specific metope, now in the Archaeological Museum at Olympia, was taken from the west end of the temple at Olympia and featured Herakles receiving the golden apples of the Hesperides from Atlas [Buitron-Oliver 1992, Plate 9, p. 96]. Athena attends Herakles, helping him hold up the skies while Atlas fetches the four golden apples from the tree of life, this being the last of the twelve labors of Herakles. Athena ruled wisdom and literature, arts and crafts, a war goddess; see [Buitron-Oliver 1992, Plate 7, pp. 92–3] for the Statue of Athena, Acropolis Museum, Athens, a marble from 480 B.C. Athena wears a peplos, a thick woolen garment belted at the waist with vertical parallel folds, right leg showing through front, left leg back. Unlike men, women were not represented nude.

The white marble of *The Eightfold Way* is open to the California sky, upheld by a serpentine prism of vertical parallel folds, echoing the traditional form of Athena in her peplos engaged in the task of helping Herakles support the heavens. One of the curves rising from the heptagon is a quotation from a fold in Athena's peplos near her neck. Plate 9 of [Buitron-Oliver 1992] is not so clear; I made a

Slide #11

Slide #15

Figure 11. Left: The 450 million year old black serpentine prism block with base end marked as a regular heptagon. The other end will be the ∞B space heptagon homotopic to this heptagon. This stone is an unremarkable grey before polishing. Right: The white marble and black serpentine are finally together in the studio after the various matching holes have been drilled for the steel rod reinforcements for stabilization during an earthquake.

sketch from the original. This quote I felt appropriate to *The Eightfold Way* in the way it involves rigid verticals emphasizing the weight of the two-, three-, and sevenfold symmetry of the tetrahedral form, a quotation from the geometrical period of those early historical times.

Robot–Stewart Platform

The precise serpentine geometry counterpoints the free marble topology in more than form, also in process, a robot water-jet for the former, a Stewart Platform computer system for the coupling. The top of the serpentine prism exactly matches one of the 24 topological hexagons carved into the surface of the tetrahedral form. How could this matching be done? This kind of matching of two stones has been done before, the Incas and the Italians each have their own tricks, we have a new one.

The abrasive Colorado River carved the Grand Canyon out of solid rock. I mused about capturing that sort of power in my studio. Looking over the south rim at the tiny filament of water glistening in the sun far below was about the

Slide #12

Figure 12. One of the three pairs of sensors of the Stewart Platform SP-2. A computer monitors the lengths of the six high tensile strength aircraft cable emerging from the semi-toroids coupled to the string potentiometers.

Slide #16

Figure 13. Regular heptagon base end of the prism being prepared to fit in the hyperbolic platform disc. The end matches the missing central heptagon of the already installed platform.

Slide #18

Slide #22

Figure 14. Left: Bill Thurston and Joe Christy attaching tapes to the spatial heptagons above for connecting strings to the regular heptagons below. Right: Close-up of the marble cloud and supporting serpentine prism with a view of the Athena quote.

filament size I saw close up in a water jet. The noise of the water jet seems to compress millions of years of erosion into a few seconds of roaring tornado sound, churning the catch chamber below into white water. This violent roar comes from a filament of water issuing from a diamond orifice under 55,000 pounds per square inch pressure. When I used a water-jet to cut the stones for *The Eightfold Way*, the water-jet was still somewhat of an experimental device. Since that time it is a common industrial tool, used to cut all manner of materials from textiles to five inch thick steel. These devices are not suitable for carving, they are through cut devices which explode from one side of the material to the other. They are also robots in the strict sense that they respond to a predetermined straight line program which allows no variation. All motions have to be calculated in advance. A complete set of 232 blocks were "dry run" out of $\frac{3}{4}$-inch plywood. This set of hyperbolic tiles was cut first and assumbled before cutting the serpentine stone. The final stone, counting the prism, was 24 black and 208 green serpentines. The Virginia green in this case was actually a bit harder stone than the black. (Greens from other parts of the country tend to be softer.) The greens tended to be the smaller stones, all the heptagons were cut to great accuracy with seven circular arc geodesic edges and seven 120-degree

Slide #19

Figure 15. Close up of some of the incision and excision boundaries of the spatial heptagonal tiling of the thetrahedroid. Note the identification tapes of Thurston and Christy.

interior angles. The cutting itself took about a week for both the debugging run with plywood and the actual cutting of the stones. One of the most difficult parts of the cutting process was how to hold the part to be cut. Part holding became more challenging for the smaller heptagons in both the wood and the stone.

By contrast to the water-jet, the Stewart Platform system in my studio, SP-2, is not a robot but an information machine [Albus et al. 1993]. The cutting tool can be moved freely to any accessible point where information is provided as to the location of a virtual image. There is no straight line program relative to the cutting process, work is interruptible at any time and point and can be continued. The operator solves all the trajectory problems as they arise, they do not need to be computed in advance.

The mating of the white and black heptagons was accomplished by the second generation of the Stewart Platform virtual image projection system, SP-2, was used in the creation of *The Eightfold Way*. SP-2 has six instead of three cables with all six lengths monitored by sensors arranged in Stewart platform format [Albus et al. 1990; 1993]. The operator interactively flies the triangle (much as if flying a helicopter). Tool tip position (x, y, z) coordinates and tool orientation (pitch, roll, yaw) are computed from the six cable lengths. Carving the *Eightfold Way* included matching two stone parts, a hand shaped heptagon in the serpentine with a matching rounded heptagon on the tetrahedral marble

Slide #27

Slide #23

Figure 16. Left: Spatial heptagon tesselation boundaries, part of a Petrie cycle, for tracing left, right, left, right, left, right, left, right and returning — or for that matter right, left, right, left, right, left, right, left and returning. Right: Matching white and black heptagons with whimsically mirror imaged pair of Helaman Ferguson signatures.

form. The SP-2 helped. First the concave heptagon was carved in the marble. This heptagon was then touched with the tip of the inactive air drill to input a cloud of points in no particular order close enough together. The three registration points were relocated in reversed order to carve the convex hand in the serpentine to hold the marble at its concave heptagon.

The SP-2 or Stewart Platform Number Two, or the VIP or Virtual Image Projection refers to one *inverse digitization* process which I have developed jointly in a CRADA (Cooperative Research and Development Agreement) between my studio and NIST (National Institute of Standards and Technology). This inverse digitization, goes from either parametric equations or a data base in the computer, into physical materials. My aesthetic choice is direct carving in the final material, e.g., subtractive carving of natural stone. The present form of this computer instrument has been strongly influenced by that aesthetic choice. The concepts are simple and powerful and can be adapted to other forms, as was the case with my series of minimal surface sculptures Costa II and Costa III.

The SP-2 itself is mathematical engineering based on a theorem of Cauchy from over a century and a half ago. Cauchy discovered many theorems referred to nowadays as Cauchy's Theorem. This one states that a convex polyhedron is

Slide #24

Figure 17. Spatial heptagons crowding together to tesselate the inside the tetrahedron.

Slide #26

Figure 18. Triple point or triskelion, this one all incision edges. This corresponds to a boundary point of the cluster. Which one?

determined if the lengths of its edges are known. Cauchy applies to the polyhedron being an octahedron of eight triangular faces, twelve edges and six vertices, the dual of a cube. The SP-2 which hangs in my studio includes two rigid equilateral triangles, one on the ceiling 13 feet on a side and one triangle suspended in midair 3 feet on a side. The other six edges are made of high tensile strength fine cable of variable length feeding under tension into six length sensors. These six lengths are then available to the computer (a MacII soon to be replaced by a G3) through an analog to digital interface. Since the six edges of the two rigid triangles are known exactly, the other six variable lengths, when known at any instant completely determine the octahedron. They determine implicitly the position and orientation of the suspended and moveable triangle, in particular the position and orientation of any tool fixed to that triangle.

A complex mathematical model originally developed for NASA for the space shuttle has been adapted for this engineering setup. The current software includes a C language implementation of this model which takes the six lengths imput and computes six coordinates which are three for the location of the tool tip and three for the orientation of the tool. This computation is done in real time on the Mac II.

It is helpful to compare the Stewart Platform system SP-2 with the traditional pointing machine. An accurate but not helpful comparison would be that a pointing machine is to the SP-2 as a hand cart is to the helicopter. Pointing machines for sculpture have been around for hundreds of years. Pointing machines, whatever their variety, refer to an existing object, a solid model or maquette, which is to be copied or enlarged. These pointing machines are slow and laborious to use, but quite effective. On the other hand, the SP-2 does not need a physical model to work from, the image can be in the computer as a data base or as equations. Digitization is a process for getting physical image coordinates into a computer data base. The SP-2 can be run in reverse, as a digitizer. A helpful description of the SP-2 is that it is an inverse digitizer. The heptagonal hand of the white marble was digitized, once the data was in in the computer, the heptagonal surface image in three dimensions was rotated around (in the computer) and then that virtual image was projected back into three dimensions, this time cut directly in the black serpentine.

Explicit quantitative sculpture includes a quantitative creation (mathematical) prior to the physical creation. The physical artifact then partakes in various ways of the original quantitative creation, but tends to be convolved with geologically or physically interesting natural materials. Technology is just emerging to make such sculpture possible in person hours instead of months or years. It should be kept in mind, *even possible at all*, due to the inhumanly huge numbers of calculations involved, once impossible, now possible.

Location

The Eightfold Way is permanently installed on the southeast patio of the Mathematical Sciences Research Institute (MSRI), at 1000 Centennial Drive, approximately 1300 feet (400 meters) above sea level (see http://www.msri.org). This land, part of the upper Berkeley hills, belongs to the University of California, although MSRI is an independent entity. Centennial Drive winds up from the Berkeley campus past the Lawrence Berkeley National Laboratory and the Lawrence Hall of Science to the Space Science Laboratory and MSRI. On the slope from the Lawrence Hall to MSRI there are parking lots; the sculpture patio, however, faces the other way, onto a fold of the hills, with a lovely view of mountainside, Oakland, and part of the San Francisco Bay (see Plate 1 and Figure 7). A wide trail, popular with joggers and walkers, leads from MSRI along a level curve of the hills; narrower trails crisscross the hillside through the scrub and scree.

Acknowledgements

I thank the former Claire Eising, William Thurston, William Reinhardt, Robert Osserman, Richard Askey, Lenore Blum, Silvio Levy, Elwyn Berlekamp, Lester Senechal, James Cannon, James Timourian, Marjorie Senechal, Stephen Wolfram, Seth Passell, Johnna Benson Cornett, Jon Ferguson, Sam Ferguson, Al Marden, Al Hales, and Al Gray.

References

[Adamowicz and Zbierski 1997] Z. Adamowicz and P. Zbierski, *Logic of mathematics: a modern course of classical logic*, Pure and Appliced Mathematics, Wiley-Interscience, 1997.

[Albus et al. 1990] J. Albus, R. Bostleman, H. Ferguson, S. Ferguson, et al., "Functional description of a String-Pot Measurement System, SP-1", Draft document, National Institute of Standards and Technology, Gaithersburg, MD, 6 April 1990.

[Albus et al. 1993] J. Albus, R. Bostleman, H. Ferguson, S. Ferguson, et al., "String-Pot Measurement System SP-1, hardware/software", Draft document, National Institute of Standards and Technology, Gaithersburg, MD, 3 March 1993.

[Apéry n.d.] F. Apéry, Slide of genus-3 form from an algebraic equation, personal communication.

[Ashley 1944] C. W. Ashley, *The Ashley Book of Knots*, Doubleday, Garden City, New York, 1944. "Every practical knot — what it looks like, who uses it, where it comes from, and how to tie it, with 7000 drawings representing over 3900 knots." See pages 9 and 671 for the wall knot, 677 for the double wall, 10 for the wall and crown, 678, 681, 683 for the Matthew Walker.

[Avery 1966] C. Avery, *Michelangelo*, Maestri della Scultura **70**, Fratelli Fabbri, Milano, 1966. See vol. 1, "La Battaglia dei Centauri", nota critica, Plates II-III (full spread, connected), XVII (cover).

[Barwise and Etchemendy 1991] J. Barwise and J. Etchemendy, "Visual information and valid reasoning", pp. 9–24 in *Visualization in teaching and learning mathematics*, edited by W. Zimmerman and S. Cunningham, MAA Notes **19**, Mathematical Association of America, Washington, DC, 1991. Reprinted as pp. 160–182 in *Philosophy and the Computer*, edited by Leslie Burkholder, Boulder, Westview Press, 1992.

[Barwise and Etchemendy 1996] J. Barwise and J. Etchemendy, "Visual information and valid reasoning", pp. 3–25 in *Logical reasoning with diagrams*, edited by G. Allwein and J. Barwise, Oxford Univ. Press, New York, 1996.

[Beck et al. 1994] J. Beck, A. Paolucci, and B. Santi, *Michelangelo: The Medici Chapel*, Thames and Hudson, London, 1994. Photographs by Aurelio Amendola. See particularly Guiliano, *Night and Day*; Lorenzo, *Dusk and Dawn.*

[Blankstein 1998] A. Blankstein, "Helaman Ferguson, Four Canoes, 1997, St. Paul, Minnesota", *Sculpture,* a publication of the International Sculpture Center (May/June 1998), 27.

[Blum et al. 1998] L. Blum, F. Cucker, M. Shub, and S. Smale, *Complexity and real computation*, Springer, New York, 1998. With a foreword by Richard M. Karp.

[Buck 1949] C. D. Buck, *A dictionary of selected synonyms in the principal Indo-European languages*, U. Chicago Press, Chicago, 1949. Reprinted in paperback, 1988.

[Buitron-Oliver 1992] D. Buitron-Oliver, *The Greek miracle: classical sculpture from the dawn of democracy, the fifth century B.C.*, National Gallery of Art, Washington, DC; distributed by Harry N. Abrams, New York, 1992. Companion book to the exhibition presented at The National Gallery of Art, Washington, DC, 22 November 1992 to 7 February 1993; The Metropolitan Museum of Art, New York, 11 March to 23 May 1993.

[Cannon 1991] J. W. Cannon, "Mathematics in marble and bronze: the sculpture of Helaman Rolfe Pratt Ferguson", *Math. Intelligencer* **13**:1 (1991), 30–39. The issue's cover shows *Alexander's Horned Sphere.* Cannon refers to Carl Faith's theory: "...it is because mathematicians are so emotional that they can become mathematicians."

[Cannon 1996] J. W. Cannon, *Math. Intelligencer* **18**:2 (1996), 73–75. A review of [Ferguson 1994].

[Chandrasekhar 1987] S. Chandrasekhar, *Truth and beauty: aesthetics and motivations in science*, U. Chicago Press, Chicago and London, 1987. See Chapter 4, "Beauty and the quest for beauty in science", the opening pages of which bring up Watson's comparison of an identity of Ramanujan to the Medici Chapel. Dick Askey brought Chandrasekhar to a small exhibition (AMS-MAA Meeting in Cincinnati, 1994) of my sculptures. C's remark was "There is a lot more to this than twisting Möbius bands out of strips of paper."

[Chumbawamba 1997] Chumbawamba, *Tubthumper* (disc), Republic, Universal Records Inc., New York, 1997. "Amnesia", track 2, 3:22-3:35, before "Drip, Drip, Drip".

[Cipra 1997] B. A. Cipra, "Quod granite demonstrandum", *SIAM News* **30**:10 (December 1997), 1.

[Cole 1998] K. C. Cole, *The universe and the teacup: the mathematics of truth and beauty*, Harcourt Brace, 1998. In Chapter 14, "Emmy and Albert", the author cites Hermann Weyl on truth vs. beauty and then has this: "Beauty in the mathematical sense is a lot more than a pretty face. It is a way of distilling the essence of things out of the messy mix that nature presents us."

[Cox et al. 1994] J. J. Cox, Y. Takezake, H. R. P. Ferguson, K. E. Kohkonen, and E. L. Mulkay, "Space-filling curves in tool path applications", *Computer Aided Design* **26**:3 (1994), 215–224.

[Coxeter 1957] H. S. M. Coxeter, "Crystal symmetry and its generalizations", *Trans. Royal Soc. Canada* (3) **51** (1957), 1–11. Contains the solution of the problem Escher was having and is the mathematical source of Escher's *Circle Limit* series.

[Coxeter 1973] H. S. M. Coxeter, *Regular polytopes*, Dover, New York, 1973.

[Coxeter 1979] H. M. S. Coxeter, "The non-Euclidean symmetry of Escher's picture 'Circle Limit III"', *Leonardo* **12** (1979), 19–25.

[Coxeter 1998] H. S. M. Coxeter, *Non-Euclidean geometry*, 6th ed., Mathematical Association of America, Washington, DC, 1998.

[Davis 1994] P. J. Davis, "Mathematics and art: the work of Helaman Ferguson", *SIAM News* **27**:5 (May/June 1994), 6–7. "A person viewing [Helaman's] pieces with no knowledge of their mathematical iconography might be compared to someone totally ignorant of western history and culture viewing a Raphael Madonna."

[Drower 1985] M. S. Drower, *Flinders Petrie: a life in archaeology*, Gollancz, London, 1985. Second edition, 1995, University of Wisconsin Press, Madison, WI. William Matthews Flinders-Petrie, a founding father of archaeology, developed field excavations (stratification). He developed measures into inductive metrology, using sequence dating methods which are now standard. He applied his system of metrology to the great pyramid at Giza and at Stonehenge. He actually spent much time penetrating the sides of the pyramid of Hawara to find the tomb of the pharoah Amenemhet III. He discovered how (other) grave robbers got through the labyrinth surrounding the tomb. An typical Egyptian pyramid has a square base and four triangular sides. The tetrahedral form we have been considering are topologically equivalent to an ordinary Egyptian pyramid with the four triangular sides tunneled into and meeting in the center (burial chamber somewhere inside). Again three cuts open up the handles and sunder the pierced solid pyramid into a ball. Petrie's books include *Inductive Metrology, or the Recovery of Ancient Measures from the Monuments*, written at age 24, *Stonehenge: Plans, Description, and Theories* (1880), and *The Formation of the Alphabet* (1912).

[Emmer 1980] M. Emmer, *The fantastic world of M. C. Escher* (VHS Video, 50 minutes), Film 7 International and M. C. Escher Foundation, 1980. Rereleased by Atlas Video, Inc., 1994. There is a sequence of H. S. M. Coxeter discussing his interaction with Escher on the Circle Limit series; Escher had done a square limit series but until he saw Coxeter's paper had not been able to do a circle limit image; I would not fault a draftsman for not discovering non-Euclidean geometry in the course of doing a woodcut.

[Escher 1982] *M. C. Escher, his life and complete graphic work*, H. N. Abrams, New York, 1982. "...with a fully illustrated catalogue ...by F. H. Bool, J. R. Kist, J. L. Locher, F. Wierda; with essays by Bruno Ernst, M. C. Escher; edited by J. L. Locher; translated from the Dutch by Tony Langham and Plym Peters." Reprinted by Abradale Press and H. N. Abrams, New York, 1992. *Circle Limit III*, woodcut unidirectional fish in five colors, yellow, green blue, green, black, diameter 415 mm, December 1959, Cat. No. 434, pp. 97, 320; *Circle Limit IV*, heaven and hell, woodcut in two colors, black and ochre, July 1960, Cat. No. 436, pp. 98, 322; *Circle Limit I*, black and white fish, Cat. No. 429, diameter 418 mm, p. 319; *Circle Limit II*, March 1959, red and black ×'s, diameter 417 mm, p. 320.

[Escher 1989] M. C. Escher, *Escher on Escher: exploring the infinite*, Harry N. Abrams, New York, 1989. With a contribution by J. W. Vermeulen; translated from the Dutch by Karin Ford. The plates for *Circle Limit I* (fishes), *Circle Limit III* (unidirectional fishes), and *Circle Limit IV* (angels and devils) can be found on pages 126, 43, and 42.

[Escher 1992] M. C. Escher, *The graphic work: introduced and explained by the artist*, Benedikt Taschen Verlag, 1992. *Circle Limit I*, white and black fish, 1958, page 22; *Circle Limit III*, five color fish, 1959, page 24; *Circle Limit IV*, white angels and black devils, 1960, page 25.

[Escher 1995] M. C. Escher, *The M. C. Escher sticker book*, Harry N. Abrams, New York, 1995. 79 imaginative stickers, including *Circle Limit III* (unidirectional fishes) and *Circle Limit IV* (angles and devils).

[Farmer and Stanford 1996] D. W. Farmer and T. B. Stanford, *Knots and surfaces: a guide to discovering mathematics*, Mathematical World **6**, Amer. Math. Soc., Providence, RI, 1996.

[Ferguson 1994] C. Ferguson, *Helaman Ferguson: mathematics in stone and bronze*, Meridian Creative Group (1-800-530-2355), Erie, PA, 1994.

[Ferguson and Forcade 1979] H. R. P. Ferguson and R. W. Forcade, "Generalization of the Euclidean algorithm for real numbers to all dimensions higher than two", *Bull. Amer. Math. Soc. (N.S.)* **1**:6 (1979), 912–914.

[Ferguson and Forcade 1982] H. R. P. Ferguson and R. W. Forcade, "Multidimensional Euclidean algorithms", *J. Reine Angew. Math.* **334** (1982), 171–181.

[Ferguson and Rockwood 1993] H. Ferguson and A. Rockwood, "Multiperiodic functions for surface design", *Comput. Aided Geom. Design* **10** (1993), 315–328.

[Ferguson et al. 1996] H. Ferguson, A. Gray, and S. Markvorsen, "Costa's minimal surface via Mathematica", *Mathematica in Education and Research* **5**:1 (1996), 5–10. The Weierstrass ζ function is used to provide an effective implementation of Costa's minimal surface.

[Ferguson et al. 1998] H. R. P. Ferguson, D. H. Bailey, and S. Arno, "Analysis of PSLQ, an integer relation finding algorithm", *Math. Comp.* (1998). To appear.

[Fischer 1986] G. Fischer, *Mathematische Modelle: aus den Sammlungen von Universitäten und Museen / Mathematical models: from the collections of universities and museums*, Vieweg, Braunschweig, 1986. Dual German-English text.

[Gell-Mann and Ne'eman 1964] M. Gell-Mann and Y. Ne'eman, *The eightfold way*, Frontiers in physics, W. A. Benjamin, New York, 1964. On pages 11–57 is reprinted the paper "The eightfold way: a theory of strong interaction symmetry", by Murray Gell-Mann, California Institute of Technology Laboratory Report CTSL-20 (1961). In this paper Gell-Mann and Ne'eman predicted quarks and a heavy particle Ω_-. Quarks were introduced theoretically by assuming three base states considered to transform according to the eight-dimensional group SU(3). One of the systems of weight points for an irreducible representation of SU(3) corresponds to an octet of baryons including the proton and neutron. Since this paper, quarks have been regarded as structural elements of baryons.

[Gray 1901] H. Gray, *Anatomy, descriptive and surgical*, 1901. Reprinted 1974 by Running Press, Philadelphia, as *Gray's Anatomy*.

[Gray 1993] A. Gray, *Modern differential geometry of curves and surfaces*, CRC Press, Boca Raton, FL, 1993.

[Greenberg 1993] M. J. Greenberg, *Euclidean and non-Euclidean geometries: Development and history*, Third ed., W. H. Freeman and Company, New York, 1993. The cover image is M. C. Escher's *Circle Limit IV*. The accompanying text describes how this was modelled on a tesellation of the hyperbolic plane by $45° - 45° - 90°$ triangles which are pairings of the alternately colored $30° - 45° - 90°$ triangles in Figure 7 of [Coxeter 1957].

[Gross and Tucker 1987] J. L. Gross and T. W. Tucker, *Topological graph theory*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley, New York, 1987.

[Hall 1986] H. T. Hall, "The transformation of graphite into diamond", *Journal of Crystal Growth* **16**:1 (March 1986), 1–8. Hall was the first to make synthetic diamonds, first with the 'Belt' press at General Electric in 16 December 1954, (U.S. Pat. Nos. 2,941,248 of June 21, 1965 and 2,947,608 of 2 Aug 1960), then with the tetrahedral press at Brigham Young University in 1956 (U.S. Pat. No. 2,918,699 of 29 December 1959). His later cubic preeses became the standard for industrial diamond manufacture. These are the diamonds I use for carving stone. Now that I think about it, it is curious that the tetrahedroid of *The Eightfold Way* is not unlike the negative space of Hall's tetrahedral press. As a sculptor I am heavily dependent upon the synthetic diamond industry of which Hall is the father. Diamonds, industrial diamonds, made synthetically are very important to me. I use an impressive array of diamond impregnated tool faces to subtract volumes of stone. Every stone worker nowadays does; these diamond tools are expensive but nowhere as expensive as they would be without synthesis. (The Tracy Hall Foundation home page is http://www.novatekonline.com/hthfound.htm; it includes his articles and patents.)

[Hill et al. 1989] S. A. Hill, P. A. Griffiths, and J. F. Bucy, "Our invisible culture: mathematics is the invisible culture of our ages", pp. 32–33 in *Everybody counts: a report to the nation on the future of mathematics education*, National Academy Press, Washington, DC, 1989.

[Holinshed 1587] R. Holinshed, *Chronicles of England, Scotland, and Ireland*, 1587. Reprinted in The Signet Classic Shakespeare Series, Sylvan Barnet, general editor; The Life of Henry V, edited by John Russel Brown, pages 172–210 (The New American Library, New York, 1965). The physical qualifications of Henry V for war as described by Holinshed include being strong and nimble, and a wrestler, leaper, and runner without compare; "in casting of great iron bars and heavy stones he excelled commonly all men".

[Homer n.d.] Homer, *The Iliad*, See, for example, the translation by Stanley Lombardo, with an introduction by Sheila Murnaghan, Hackett Pub. Co., Indianapolis, 1997. The lines referred to are summarized as follows: Patroclus is consumed with Achilles' refusal to fight after his squabble with Agamemnon over the prize girl Briseis. Patroclus decides to enter the fray with the Trojans in Achilles' stead. The Trojan prince Hector and his charioteer Cebriones drove straight through the other Greeks to get to Patroclus, who scooped up a jagged piece of granite with one hand and threw it at Cebriones, smashing his brows together and shattering his skull so his eyeballs spurted out. In final convulsion Cebriones flipped backwards out of his chariot. Nice dive, comments Patroclus. Rocks picked up off the ground and thrown figure in several episodes in the Iliad.

[Hughes 1997] R. Hughes, *American visions: the epic history of art in America*, Alfred A. Knopf, New York, 1997. On page 574, he points out that for the Puritan the word was law and the image was a delusion and snare. Of extreme puritan heritage myself, emphasis on ex now, I myself responded to the linear word made three-dimensional flesh.

[Hurwitz 1893] A. Hurwitz, "Über algebraische Gebilde mit eindeutigen Transformationen in sich", *Math. Annalen* **41** (1893), 403–442.

[Jahnke and Emde 1945] E. Jahnke and F. Emde, *Tables of functions with formulae and curves*, 4th ed., Dover, New York, 1945. See Addendum VI, "The exponential function $e^{1/z}$", p. 39.

[Kitagawa 1971] J. M. Kitagawa, "The Eightfold Path to Nirvana", pp. pages 88–101 in *Great religions of the world*, edited by M. Severy, National Geographics Society, 1971. The expression is aṣṭāṅgika-marga in Sanskrit and aṭṭhaṅgika-magga in the more recent Pali. The Eightfold Path is the way shown by the Four Noble Truths summarized in practice by the Three Learnings. The eight unfolded ways briefly summarized are right understanding (view), right thought, right speech, right action (conduct), right livelihood, right effort, right mindfulness, right concentration (meditation). By following the Eightfold Way one is offered the hope of being reborn more suitably ultimately achieving nīrvāṇa. Perhaps for the context of the sculpture, one can think of the classic eight as right understanding, left thought, right speech, left action, right livelihood, left effort, right mindfulness, left concentration as one recites around the sculpture. The adjective *right* connotes that they are all in accord with reality as it is (yathā-bhūtaṃ). Left here may be more in accord with "reality as it is" than a right turn, especially if one is to appreciate the underlying mathematical reality of the symmetry. To understand the eightfold way you must follow it as in ἔρχου καὶ ἴδε or "come and see".

[Lehmer 1933] D. H. Lehmer, "Factorization of certain cyclotomic functions", *Ann. Math.* (2) **34** (1933), 461–479.

[Levenson 1996] C. B. Levenson, *Symbols of Tibetan Buddhism*, Éditions Assouline, Paris, 1996. Photographs by Laziz Namani; Translated by Nessim Marshall. See particularly Chapter 3, "The wheel of the law".

[Levy 1994] S. Levy, "Mathematics in stone and bronze", *The Mathematica Journal* **4**:3 (1994), 14–16. A review of [Ferguson 1994].

[Lichtenberg 1978] D. B. Lichtenberg, *Unitary symmetry and elementary particles*, 2nd ed., Academic Press, 1978. Chapter 9, "The eightfold way", describes the situation: Strongly (not electromagnetic) interacting elementary particles are called hadrons. The dynamics of elementary particle interaction (scattering) is complex, so even approximate symmetries are vital to making predictions. An acceptable aprroximation is on the order of the fine structure constant ($\sim 1/137$). Unitary transformations are associated with conservation laws, and the matrix groups SU(2), SU(3) and SU(4) all provide approximate symmetries. The "eight" in "eightfold way" comes from the sequence beginning with $1, 8, 10, 27$ where the 8 is the most frequent, higher mass of the vector mesons or baryons. 1 is trivial, 10 and 27 refer to higher excited mass (only 10 has been seen experimentally so far). Note that the Clebsch–Gordan series for the Kronecker product of an eight-dimensional zero-triality representation of SU(3) is $\mathbf{8} \otimes \mathbf{8} = \mathbf{1} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{10} \oplus \overline{\mathbf{10}} \oplus \mathbf{27}$; thus $8^2 = 1 + 2 \cdot 8 + 2 \cdot 10 + 27$. Baryon number B and charm C are defined for hadrons. Baryons have $B = 1$, mesons have $B = 0$; baryons are composed of three quarks, mesons of a quark-antiquark pair. Both form a charm octet, but not quite a mass octet. The eight baryons are denoted by $(p, n, \Lambda, \Sigma^+, \Sigma^-, \Sigma^0, \Xi^0, \Xi^-)$ with masses (in MeV, with $c = 1$) $(938, 940, 1116, 1189, 1192, 1314, 1321)$, with isospin I $(\frac{1}{2}, \frac{1}{2}, 0, 1, 1, 1, \frac{1}{2}, \frac{1}{2})$, giving one isospin singlet, one isospin triplet and two isospin doublets; with third component of isospin I_3 $(\frac{1}{2}, -\frac{1}{2}, 0, 1, 0, -1, \frac{1}{2}, -\frac{1}{2})$, with hypercharge Y $(1, 1, 0, 0, 0, 0, -1, -1)$ where the last six $(Y \neq 1)$ are called hyperons. The isospin I_3 and hypercharge Y together give the hexagonal weight diagram. The eight mesons are denoted by (π^0, π^+, π^-, K^+, K^-, K^0, $\bar{K}^0$, η), with masses (135, 140, 140, 494, 494, 498, 498, 549); there is an associated singlet η' with mass 957. Mesons are messier than baryons for a number of reasons; note the symmetry broken masses relative to the charm octet of $(0, 0, 0, 0, 0, 0, 0, 0)$.

[Luke 1977] Y. L. Luke, *Algorithms for the computation of mathematical functions*, Academic Press, New York, 1977.

[MacLane 1971] S. MacLane, *Categories for the working mathematician*, Graduate Texts in Mathematics **5**, Springer, New York, 1971.

[Morrison 1988] P. Morrison, "On broken symmetries", pp. 54–70 in *On aesthetics in science*, edited by J. Wechsler, Birkhäuser, Boston, 1988.

[Nassau 1980] K. Nassau, *Gems made by man*, Chilton Book Co., Radnor, PA, 1980.

[Netter 1996] F. H. Netter, *Atlas of human anatomy*, 2nd ed., Ciba-Geigy Corporation, Summit, NJ, 1996. See particularly Plate 4 for bregma (anterior) and lambda (posterior); Plate 2 for other triple points and pterion; and Plate 48.

[Neugebauer 1975] O. Neugebauer, *A history of ancient mathematical astronomy*, Studies in the history of mathematics and physical sciences **1**, Springer, Berlin, 1975.

[Newman 1956] J. R. Newman, *The world of mathematics*, Simon and Schuster, New York, 1956.

[Nomachi 1997] K. Nomachi, *Tibet*, Shambhala Publishers, Boston, 1997.

[Pappas 1994] T. Pappas, *The magic of mathematics*, World Wide Publishing/Tetra, 1994. This is the only source I have seen for visual counting systems such as the Chinese rods.

[Poeschke 1996] J. Poeschke, *Michelangelo and his world: sculpture of the Italian Renaissance*, Harry N. Abrams, New York, 1996. Photographs by Albert Hirmer and Irmgard Ernstmeier-Hirmer. Plates 3, 4, 5, 6 show Michelangelo's *The Battle of the Centaurs*, c. 1490–1492, Casa Buonarroti, Florence.

[Powell 1995] A. Powell, *Living Buddhism*, University of California Press, Berkeley, 1995. Photography by Graham Harrison. See particularly p. 125, the facade on a Tibetan temple (gompa) near Dharmsala.

[Ratcliffe 1994] J. G. Ratcliffe, *Foundations of hyperbolic manifolds*, Graduate Texts in Math. **149**, Springer, New York, 1994.

[Reid 1991] A. W. Reid, "Arithmeticity of knot complements", *J. London Math. Soc.* (2) **43**:1 (1991), 171–184.

[Richer 1890] P. M. L. P. Richer, *Anatomie artistique: description des formes extérieures du corps humain au repos et dans les principaux mouvements*, Plon, Paris, 1890. Translated by Robert Beverly Hale as *Artistic anatomy*, Watson-Gupthill, New York, 1971.

[Riley 1975] R. Riley, "A quadratic parabolic group", *Math. Proc. Cambridge Philos. Soc.* **77** (1975), 281–288.

[Schensted 1976] I. V. Schensted, *A course on the application of group theory to quantum mechanics*, NEO Press, Peaks Island, ME, 1976. Section 6.4: Applications to some physical problems. Spin functions again. Something about the Gell-mann Ne'eman elementary particle scheme.

[Seifert 1934] H. Seifert, "Über des Geschlecht von Knoten", *Math. Annalen* **110** (1934), 571–592.

[Senechal 1996] M. Senechal, *Quasicrystals and Geometry*, paperback ed., Cambridge University Press, New York, 1996. The frontispiece shows the sculpture "Aperiodic Penrose Torus, α".

[Sequin n.d.] C. Sequin, "TetraThing, computer generation of genus-3 tetrahedroids", personal communication (letter and photographs).

[Simmons 1991] G. L. Simmons, *Differential equations with applications and historical notes*, 2nd ed., McGraw-Hill, New York, 1991. From page xix: "Mathematics is one of the supreme arts of civilization; like the other arts, it derives its grandeur from the fact of being a human creation."

[Snow 1959] C. P. Snow, *The two cultures and the scientific revolution*, Cambridge University Press, New York, 1959.

[Spenceley et al. 1952] G. W. Spenceley, R. M. Spenceley, and E. R. Epperson, *Smithsonian logarithmic tables to base e and base 10*, Smithsonian Institution, Washington, 1952. Reprinted by The Lord Baltimore Press, Baltimore, 1960.

[Thurston 1997] W. P. Thurston, *Three-dimensional geometry and topology, vol. 1*, edited by S. Levy, Princeton Mathematical Series **35**, Princeton University Press, Princeton, NJ, 1997.

[Whittaker and Watson 1927] E. T. Whittaker and G. N. Watson, *A course of modern analysis; an introduction to the general theory of infinite processes and of analytic functions; with an account of the principal transcendental functions*, 4th ed., The University Press, Cambridge, 1927. Reprinted in 1963.

Helaman and Claire Ferguson
helamanf@access.digex.net
cferguso@sophia.smith.edu
10512 Pilla Terra Court
Laurel, MD 20723-5728
United States

The Eightfold Way
MSRI Publications
Volume **35**, 1998

Invariants of $\mathrm{SL}_2(\mathbb{F}_q) \cdot \mathrm{Aut}(\mathbb{F}_q)$ Acting on $\mathbb{C}^n$ for $q = 2n \pm 1$

ALLAN ADLER

In fond memory of my friend and teacher Michio Kuga.

Abstract. We define bicycles and present the Bicycle Conjecture, which is false in general but which we believe is neverthelesss quite useful, and derive from it specific open conjectures about some explicit conjectural generators of the bicycles of invariants of components of the Weil representation of $\mathrm{SL}_2(\mathbb{F}_q) \cdot \mathrm{Aut}(\mathbb{F}_q)$ and $\mathrm{Sp}_{2r}(\mathbb{F}_p)$. Construction of these generators depends on our result that the Weil representation has a unique invariant 3-tensor and our explicit computation of it, and on results on intertwining operators given in an appendix. We then give a tentative definition of the notion of "geometric construction" based on covariants. In spite of its limited scope, it is adequate for the purposes of this article. We prove that the modular curve $X(p)$ can be constructed geometrically from that 3-tensor provided p is a prime ≥ 11 and $\neq 13$. This uses our determination of the automorphism group of the invariant 3-tensor. The conjectural generators for the bicycle of invariants of $\mathrm{SL}_2(\mathbb{F}_q) \cdot \mathrm{Aut}(\mathbb{F}_q)$ are inspired by and generalize the generators given in the Klein–Fricke treatise for the ring of invariants of the three-dimensional representation of $\mathrm{PSL}_2(\mathbb{F}_7)$. That includes, in particular, the quartic invariant defining the Klein curve.

1. Introduction

The work described in this article was motivated by a desire to understand from a general point of view the results of Felix Klein on the equations defining modular curves of prime order, especially his remarkable discovery that the modular curve $X(11)$ is the singular locus of the Hessian of the cubic threefold

$$v^2w + w^2x + x^2y + y^2z + z^2v = 0.$$

1991 *Mathematics Subject Classification.* 15A72, 08A02, 08A40, 20C15, 20C30, 20G20, 12H05, 13N10, 13P99, 14A99, 14N99, 17D99, 20C33.

Key words and phrases. Bicycle, Galois group, hessian, invariant differential operator, pfaffian, ring of invariants, self-adjoint group, Weil representation.

This same desire has motivated much of my work over the years (see references in the bibliography), including the computation in [Adler 1981; 1992b] of the ring of invariants of a five-dimensional complex representation of $\mathrm{PSL}_2(\mathbb{F}_{11})$ and the joint work [Adler and Ramanan 1996] on moduli of abelian varieties. At the same time, these efforts have led to other problems of interest in their own right.

In this paper, we make our first attempt at a synthesis of what we have learned from our efforts. We begin in Section 2 with some general considerations about rings of invariants introduced in [Adler 1981; 1992b], specifically the concept of a *bicycle*. A bicycle is a ring equipped with an additional structure of left module over itself. The ring of invariants of a self-adjoint group of operators or, more generally, of a weakly self-adjoint group, as in Definition 2.1, is an example of a bicycle. This fact enables one to generate rings of invariants from a small number of generators using bicycle operations. After introducing the notion of a bicycle, we then state a general conjecture (2.4), called The Bicycle Conjecture, about the bicycle of invariants of a finite group. As an example in Section 2.7 shows, the conjecture is false in general. Nevertheless, we believe that it provides a powerful tool for computing rings of invariants. The papers [Adler 1981; 1992b], show how this can work.

One weakness with the Bicycle Conjecture is that it requires one to begin with some already computed invariants. Producing explicit invariants can often be quite difficult by direct computations. Therefore it is useful to know of families of representations of finite groups for which one can produce such invariants by pure thought. We begin in Section 3 with a brief discussion of the invariants of a complex three-dimensional representation of $\mathrm{SL}_2(\mathbb{F}_7)$. These were computed by Klein [1879a] and his ingenious construction of proposed generators in that case already exhibits many features of the general case. Indeed, by adapting the tricks Klein originally used, much of the work is already done for us. In Section 4, we refer to the results of [Adler 1992a; 1994], in which an explicit invariant 3-tensor Θ was constructed for the Weil representation of $\mathrm{SL}_2(\mathbb{F}_q)$, where $q = p^r$ is an odd prime power. Starting with this 3-tensor, one can construct other invariants by considering its covariants [Dieudonné and Carrell 1971; Grace and Young 1903], when it happens to be symmetric, and also by considering certain intertwining operators of the second tensor power of the Weil representation of finite symplectic groups. Thus, we have some explicit invariants and we specialize the Bicycle Conjecture to the case of these invariants. The result is then a very specific conjecture (4.2), called the Θ Conjecture, regarding the generators of the bicycle of invariants of the $\mathrm{SL}_2(\mathbb{F}_q) \cdot \mathrm{Aut}(\mathbb{F}_q)$ in the component of its Weil representation of dimension $(q + \varepsilon)/2$, where ε is the quadratic character of -1 in $\mathbb{F}_q$. We do not know of any explicit invariants of $\mathrm{SL}_2(\mathbb{F}_q)$ which are not invariants of the larger group $\mathrm{SL}_2(\mathbb{F}_q) \cdot \mathrm{Aut}(\mathbb{F}_q)$.

In Section 5, we present a similar conjecture regarding the ring of invariants of the finite symplectic group $\mathrm{Sp}_{2r}(\mathbb{F}_p)$ when -2 is a quadratic residue modulo p. In this case, we can give an explicit quartic invariant Ω for the group as well as certain invariants which we express as explicit covariants of Ω. By applying the Bicycle Conjecture to these invariants, we obtain conjectural generators (Conjecture 5.2) for the bicycle of invariants of $\mathrm{Sp}_{2r}(\mathbb{F}_p)$ (or at least those of even degree) on the component of the Weil representation of dimension $(p+1)/2$. This conjecture is called the Ω Conjecture. From it, we can deduce other specific conjectures.

Having shown the utility of covariants in formulating explicit conjectures regarding the generators of bicycles of invariants, it is natural to ask how powerful a tool covariants provide. More precisely, given one invariant f of a finite group G, which invariants of G arise as covariants of f? Thanks to the excellent help of Gerry Schwarz (Theorem 7.1) and of David Vogan (Theorem 7.3), we have some answers to such questions and we present them in Corollary Theorem 7.5, Corollary 7.6 and Theorem 7.7.

In Section 6, we draw attention to some of the philosophical implications of questions and results of this type. More precisely, if we follow Klein in describing geometry as that which is preserved by a group action, then we have the right to ask: if that is what we mean by geometry, what do we mean by a geometric construction? For the case of classical projective geometry, we tentatively define the notion of geometric construction in terms of covariants. Very likely, ours is not the best definition and we give some criticisms of it as well in Section 6. However it does serve our purposes in this paper. These considerations also allow us to compare the geometry imposed on complex projective space by G with classical projective geometry.

As a result of the concepts introduced in Section 6 and the results of Gerry Schwarz and David Vogan mentioned above, we are able to give a qualitative generalization of the theorem of Klein about $X(11)$ mentioned in the first paragraph of this section: we prove (Theorem 7.8) that when p is a prime ≥ 11 and $\neq 13$ there *exists* a geometric construction of the modular curve $X(p)$ from the invariant 3-tensor Θ. More precisely, if -1 is a square modulo p, then one can construct Klein's A-curve of level p from the restriction of Θ to the even part of the Weil representation of $\mathrm{SL}_2(\mathbb{F}_p)$, while if -1 is not a square modulo p then one can construct Klein's z-curve from the restriction of Θ to the odd part of the Weil representation. Clearly, one cannot formulate such a theorem without asking what one means by a geometric construction.

In view of the importance of the Weil representation in our work, we include an appendix (Section 8) describing the Weil representation and the fundamental intertwining operators used in Sections 4–7. This appendix may be regarded as a sequel to [Adler 1989].

2. Group Representations and Bicycles

Let k be a field and let V be a finite-dimensional vector space over k. The k-linear functions from V to k form a vector space which we denote V^* and which we call the dual space of V. There is a natural pairing $[\,\cdot\,,\cdot\,] : V \times V^* \to k$ defined by evaluation of elements of V^* at points of V, that is, by the rule

$$[v, v^*] = v^*(v)$$

for all $v \in V$ and all $v^* \in V^*$. We will write k-linear operators on V as operators on the left and k-linear operators on V^* on the right. If α is an endomorphism of the vector space V, there is one and only one endomorphism α^* such that

$$[\alpha v, v^*] = [v, v^* \alpha^*] \tag{2.1}$$

for all $v \in V$ and all $v^* \in V^*$. Denote by $S[V]$ the symmetric algebra on V and by $S[V^*]$ the symmetric algebra on V^*. Every invertible k-linear transformation α of V extends uniquely to an automorphism of the graded k-algebra $S[V]$. We will denote that automorphism $S(\alpha)$. Similarly, we will denote by $S(\alpha^*)$ the unique extension of α^* to an automorphism of the graded k-algebra $S[V^*]$. As in the case of operators on V^*, the operator $S(\alpha^*)$ will be written on the right. If v^* is any element of V^*, the mapping $v^* : V \to k$ extends uniquely to a derivation D_v of the symmetric algebra $S[V]$. Furthermore, if α is any invertible k-linear transformation of V, the identity (2.1) implies that

$$D_{v^*\alpha^*} = D_{v^*} \circ S(\alpha). \tag{2.2}$$

Furthermore, the operators D_{v^*} with $v^* \in V^*$ commute with each other and generate the algebra, denoted $\mathcal{D}(V)$, of differential operators with constant coefficients of $S[V]$. The mapping $v^* \mapsto D_{v^*}$ extends uniquely to an isomorphism D of $S[V^*]$ onto $\mathcal{D}(V)$. The image of an element f^* of $S[V^*]$ under D will be denoted D_{f^*}. The identity (2.2) then extends to the identity

$$D_{f^* S(\alpha^*)} = D_{f^*} \circ S(\alpha).$$

Let ρ be a representation of a group G as invertible k-linear transformations on V. Then the dual representation ρ^* of G on V^* is defined by $\rho^*(g)v^* = v^*\rho(g^{-1})^*$. In particular, V^* is a left G-module with respect to ρ^*. Let σ be an automorphism of G. Then the composition $\rho \circ \sigma$ of ρ with σ is also a representation of G on V. We denote that representation by ρ^σ.

Definition 2.1. By a *weakly self-adjoint representation* we will mean a quintuple $(G, \rho, \sigma, \tau, \phi)$ where G is a group, ρ is a representation of G on a vector space V over k, σ is an automorphism of G, τ is an automorphism of k and $\phi : V \to V^*$ is a τ-semilinear isomorphism of the vector space V onto its dual space V^* such that

$$\phi(\rho^\sigma(g)v) = \rho^*(g)\phi(v) \tag{2.3}$$

for all $g \in G$ and all $v \in V$, where ρ^* denotes the dual representation of ρ and ρ^σ denotes the representation $\rho \circ \sigma$. Thus, ϕ is a τ-semilinear intertwining operator between ρ^* and ρ^σ. When it is not necessary to specify ρ, σ, τ and ϕ, we will sometimes simply speak of G as being a *weakly self-adjoint group*.

Suppose $(G, \rho, \sigma, \tau, \phi)$ is a weakly self-adjoint representation. The isomorphism ϕ extends uniquely to a τ-semilinear isomorphism $S(\phi)$ of the k-algebra $S[V]$ onto the k-algebra $S[V^*]$. The identity (2.3) then implies that

$$\phi(S(\rho^\sigma(g)f)) = S(\rho^*(g))S(\phi(f))$$

for all $f \in S[V]$. Composing the τ-semilinear algebra isomorphism $S(\phi)$ with the algebra isomorphism D, we obtain the τ-semilinear algebra isomorphism $D \circ S(\phi)$, which we will denote D^ϕ. If f, p are elements of $S[V]$ we denote by $f \#_\phi p$ the result of applying the differential operator $D^\phi(f)$ to the element p of $S[V]$. We then have

$$(S(\rho(g))f) \#_\phi (S(\rho(g))p) = S(\rho(g))(f \#_\phi p)$$

for all $f, p \in S[V]$ and all $g \in G$.

In particular, we have the following two results.

Proposition 2.2. *Let $G, \rho, V, V^*, \phi, D^\phi$ be as above. Suppose that f is an element of $S[V]$ invariant under the representation ρ of G. Then $D^\phi(f)$ is a differential operator on $S[V]$ commuting with the operators $\rho(g)$ for all $g \in G$.*

Proposition 2.3. *Let $G, \rho, V, V^*, \phi, D^\phi$ be as above. Suppose that f, p are elements of $S[V]$ invariant under the representation ρ of G. Then $f \#_\phi p$ is also an element of $S[V]$ invariant under the representation ρ.*

Thus, in the situation we are considering, the ring $S[V]^G$ of invariants for the representation ρ is closed under the operation $\#_\phi$. Using it, one can often represent invariant elements of $S[V]$ with considerable brevity. It also offers the advantage that from a very small number of invariants, one can generate the entire ring of invariants by means of the new operation $\#_\phi$ on invariant polynomials. For example [Adler 1981; 1992b], in the case of the group $\mathrm{PSL}_2(\mathbb{F}_{11})$ in an irreducible representation of degree 5 over the field of complex numbers, the transcendence degree of the ring of invariants over the field of complex numbers is 5 but we are able to generate it from an invariant of degree 3 and an invariant of degree 5 using ring operations and the new operation $\#_\phi$.

It therefore seems appropriate to begin the study of a new type of algebraic structure consisting of a ring R and a homomorphism from R into the ring of endomorphisms of the additive group of R. Thus, R is a ring with an exotic structure of left module over itself. That module structure is a ring homomorphism from the ring R into the ring of endomorphisms of the additive group of R. I call such a structure a *bicycle*. Therefore, we have associated a bicycle to the quintuple $(G, \rho, \sigma, \tau, \phi)$ which we will call the *bicycle of invariants* of G acting on V. This bicycle does depend on σ, τ and ϕ as well, but in practice these will be known from the context and we omit explicit mention to avoid circumlocution. In the case of the bicycle of invariants, the exotic module structure is simply D^ϕ. Hence, we may denote the bicycle of invariants by $(S[V]^G, D^\phi)$.

The category of rings is naturally embedded in the category of bicycles via the regular representations. What we have in the case of bicycles of invariants is a class of examples of of bicycles which do not arise in this way. This class has other special features which ought to be noted. First, in the bicycle of invariants of $(G, \rho, \sigma, \tau, \phi)$ the exotic module structure is an action of the ring on itself by differential operators. Thus, it is appropriate to speak of it as a *differential bicycle.*

The notion of differential bicycle is quite general, since one has a notion of differential operator on any commutative ring with unity: a differential operator of order 0 on such a ring R is just multiplication by an element of R while, for $n > 0$, an endomorphism of the additive group of R is a differential operator of order $\leq n$ if its commutator with every differential operator of order 0 is a differential operator of order $< n$. In particular, it makes sense to speak of the degree of such a differential operator as being the smallest integer n for which the operator has degree $\leq n$. This defines a filtration of the ring $\mathrm{Diff}(R)$ of differential operators on R but in general not a grading.

If the ring R happens to have a grading, one can speak of a different notion of degree for a differential operator, which we will call the *graded degree* of the operator. We will say that a differential operator $\mathcal{D}$ has graded degree n if, viewed as an endomorphism of the additive group of R, which is a graded abelian group, $\mathcal{D}$ has degree $-n$. It is not necessarily the case that a differential operator on a ring with a grading has a graded degree. Nor is it necessarily the case in general that the graded degree coincides with the degree of the differential operator in case the graded degree is well defined. If a differential operator on a graded ring is such that its graded degree is well defined and equals the degree of the differential operator, we will say that the operator has *good grades.*

If $S[V]^G$ is the ring of invariants and if we denote by M the underlying additive group of $S[V]^G$ with its exotic left module structure D^ϕ, then then we can view M as a graded module by defining the grade in M of a form of degree d to be $-d$. Hence, we introduce the concept of a *graded bicycle* by saying that a graded bicycle is a bicycle (S, Φ) such that S is a graded ring and such that whenever x and y are elements of S homogeneous of degrees m and n respectively the element $\Phi(x)(y)$ of S is homogenous of degree $n - m$. Thus the ring of invariants of a weakly self-adjoint group is a graded bicycle. Furthermore, if (S, Φ) is a graded bicycle and also a differential bicycle, we will say that (S, Φ) is a *differential graded bicycle.* We do *not* assume that for every homogeneous element x of S, the differential operator $\Phi(x)$ has good grades. The ring of invariants of a weakly self-adjoint group is a differential graded bicycle.

Although the formal definition of a bicycle as an algebraic structure is new, the practice of converting invariants into invariant differential operators goes back to roughly the middle of the 19th century. For example, in the classical study of invariants of binary forms, one in effect uses the fact that the natural representation of $\mathrm{SL}_2(\mathbb{C})$ on $\mathbb{C}^2$ is symplectic and gives rise to a bicycle structure on the invariants of binary forms of degree n.

(Actually, the structure is richer in this case than just the bicycle structure. One has, for example, transvection operators $(f,g)^k$ for every nonnegative integer k and the bicycle operation $f\#g$ is proportional to $(f,g)^s$, where s is the degree of f.)

In the case of the ring of invariants for the simple group of order 660 in a five-dimensional irreducible representation, we have given generators and relations for the ring ([Adler 1981; 1992b]; see also Section 6 of this paper). But it would be interesting to know how to give a presentation of the *bicycle* of invariants.

In connection with the bicycle of invariants of $(G,\rho,\sigma,\tau,\phi)$, we may also consider the following rings:

(1) the ring $\mathcal{D}_1$ of differential operators generated by $S[V]^G$ and $D^\phi(S[V]^G)$;
(2) the ring $\mathcal{D}_2=(\mathrm{Diff}(S[V]))^G$ of G invariant polynomial differential operators on V;
(3) the ring $\mathcal{D}_3=\mathrm{Diff}(S[V]^G)$ of differential operators on $S[V]^G$;
(4) the ring $\mathcal{D}_4$ of all differential operators on the quotient field of $S[V]$ which leave $S[V]^G$ invariant modulo those that annihilate it.

It would be interesting to understand the relation among these four rings in more detail. For example, in general $\mathcal{D}_1$ is not equal to $\mathcal{D}_2$, as shown by the following counterexample ([Levasseur and Stafford 1995], after the proof of Theorem 5): one lets G be a cyclic group of order 3 acting nontrivially on $V=k=\mathbb{C}$ by multiplication by cube roots of unity. On the other hand, one does have $\mathcal{D}_1=\mathcal{D}_2$ in case G is a Weyl group acting by reflections [Levasseur and Stafford 1995, Theorem 5; Wallach 1993]. (I am indebted to David Vogan for bringing the results of these two papers to my attention.)

In view of the ease with which the bicycle structure cuts across the lines usually drawn by algebraic independence, it is tempting to make the following conjecture:

THE BICYCLE CONJECTURE 2.4. *Let $(S[V]^G,\#^\phi)$ be the bicycle of invariants of $(G,\rho,\sigma,\tau,\phi)$. Let $P_1,\dots,P_r$ be homogeneous elements of $S[V]^G$. Assume that the intersection of the automorphism groups of $P_1,\dots,P_r$ is equal to G. Then every element of $S[V]^G$ can be obtained from $P_1,\dots,P_r$ using ring operations, scalar multiplication and the new operation $\#_\phi$.*

REMARK 2.5. Let m be the greatest common divisor of the degrees of $P_1,\dots,P_r$. Let M be a multiple of m Let H denote the cartesian product of the group G and the group of M-th roots of unity. We can then extend the quintuple $(G,\rho,\sigma,\tau,\phi)$ to a quintuple $(H,\rho',\sigma',\tau,\phi)$ where ρ' sends an M-th roots of unity ξ to scalar multiplication by ξ and where σ' is the identity on M-th roots of unity. We can then apply the Bicycle Conjecture to this extended quintuple. Let $Q_1,\dots Q_s$ be homogeneous elements of $S[V]^G$ the intersection of whose automorphism groups is H. Then every invariant of G whose degree is divisible by M is obtained from $Q_1,\dots,Q_s$ using ring operations, scalar multiplication and the new operation

$\#_\phi$. This follows at once from the Bicycle Conjecture and from the observation that an invariant of G has degree divisible by M if and only if it is an invariant of H.

REMARK 2.6. One example of the Bicycle Conjecture would be the assertion that the bicycle of invariants of the Monster in its faithful irreducible represention of lowest degree is generated by the invariant quadratic form and Griess' invariant cubic form.

REMARK 2.7. It is necessary to make some requirement on the automorphism groups of the forms $P_1, \ldots, P_r$. Without it, one can easily obtain counterexamples. For example, let G be the trivial group and let the set of P_i be empty. One can also take G to be the trivial group acting on a one-dimensional complex vector space, letting $r = 1$ and $P_1 = x^2$. Finally, one can take G to be any subgroup of the symmetric group S_n other than S_n itself and consider the representation of degree n of G given by permuation of the coordinates of $\mathbb{C}^n$. One can then take $P_1, \ldots, P_n$ to be the elementary symmetric functions of $x_1, \ldots, x_n$ and let the bicycle structure be given by

$$f\#g = f\Big(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\Big)g.$$

The polynomials $P_1, \ldots, P_n$ are invariant under S_n and so will all polynomials derived from them by bicycle operations, so one won't get all of the invariants of G in this way.

As stated, the hypothesis of the Bicycle Conjecture is too weak. The conjecture is false for the natural permutation representation of the symmetric group S_n. Indeed, if for $k \geq 1$ we denote by α_k the sum of the k-th powers of the n variables x_i, $1 \leq i \leq n$, then the bicycle generated by α_i, α_j, with $\gcd(i, j) = 1$, $i < j$, $j > 2$, is the polynomial ring generated by all α_k with $1 \leq k \leq j$. In particular, it doesn't contain α_n if $j < n$.

One could strengthen the hypotheses by requiring that at least one or even that all of the P_i have automorphism group G. It might also be that one needs to assume the representation ρ is irreducible. One could also require that the degrees of the P_i be greater than or equal to some lower bound. Finally, whatever their degrees, one could claim only that the conjecture be true for generic choices of the P_i's. In the absence of *any* nontrivial example of a quintuple $(G, \rho, \sigma, \tau, \phi)$ for which one can prove the Bicycle Conjecture for every choice of $P_1, \ldots, P_r$ satisfying even the strictest conditions we might wish to impose, it is pointless to make the conjecture more precise at this point. However, for definiteness, we will retain the version stated above throughout this paper. (It is reasonable to expect one of the variants of the conjecture mentioned here to hold and to expect that the conjectural generators we propose here for the bicycles of generators of $\mathrm{SL}_2(\mathbb{F}_q) \cdot \mathrm{Aut}(\mathbb{F}_q)$ do in fact generate.)

PROBLEM 2.8. In view of the detailed knowledge we have about symmetric polynomials, and more generally about Weyl group invariants, it seems plausible that one could actually prove the Bicycle Conjecture if the polynomials P_i are chosen to have sufficiently high degree (e.g. at least one of them $> n$ in the case of S_n) and to be generic.

In order to provide further tests of the Bicycle Conjecture, in Section 4 we will present a more precise conjecture for the bicycle of invariants of the irreducible representations of degree $(q \pm 1)/2$ of $\mathrm{SL}_2(\mathbb{F}_q) \cdot \mathrm{Aut}(\mathbb{F}_q)$.

3. The Tricks of Felix Klein for $\mathrm{PSL}_2(\mathbb{F}_7)$

In this section we present the generators discovered by Felix Klein for the ring of invariants of a three-dimensional complex representation of $\mathrm{PSL}_2(\mathbb{F}_7)$. Our reason for presenting this separately is that we will find that Klein's tricks, supplemented with some of our own, suffice to describe conjectural generators for $\mathrm{SL}_2(\mathbb{F}_q)$ in one component of the Weil representation in general.

The first invariant discovered by Klein is the quartic $x^3y + y^3z + z^3x$, which we will denote f in this section, following Klein. Klein was motivated to find an invariant of this degree because he knew that it would be the equation defining an embedding of the modular curve $X(7)$ of level 7 as a plane quartic curve.

The second invariant found by Klein is the Hessian of f, which he divided by a superfluous constant and denoted ∇. Explicitly,

$$\nabla = \frac{1}{54} \begin{vmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{vmatrix} = 5x^2y^2z^2 - xy^5 - yz^5 - zx^5.$$

Thus, one way to get a new invariant is to compute the Hessian of a known invariant. To get the invariant C of degree 14, he bordered the Hessian matrix with the partials of the Hessian and took the determinant, dividing by a numerical factor:

$$C = \frac{1}{9} \begin{vmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} & \frac{\partial \nabla}{\partial x} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} & \frac{\partial \nabla}{\partial y} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} & \frac{\partial \nabla}{\partial z} \\ \frac{\partial \nabla}{\partial x} & \frac{\partial \nabla}{\partial y} & \frac{\partial \nabla}{\partial z} & 0 \end{vmatrix}.$$

Thus, another trick to obtain a new invariant from an old one is to border the Hessian with the first partials of the Hessian. Finally, there is Klein's trick of taking the Jacobian of the 3 algebraically independent forms f, ∇ and C to produce the invariant K of degree 21:

$$K = \text{const} \cdot \begin{vmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial \nabla}{\partial x} & \dfrac{\partial C}{\partial x} \\ \dfrac{\partial f}{\partial y} & \dfrac{\partial \nabla}{\partial y} & \dfrac{\partial C}{\partial y} \\ \dfrac{\partial f}{\partial z} & \dfrac{\partial \nabla}{\partial z} & \dfrac{\partial C}{\partial z} \end{vmatrix}.$$

Thus, one can always try to produce new invariants from old ones by taking Jacobian determinants. Note that whereas f, ∇ and C all have even degree, the invariant K has odd degree. In general, if one has n algebraically independent forms of even degree in n variables, one can take their Jacobian and obtain a nonzero form. Moreover, if n is odd, as it is in our case, the Jacobian will have odd degree. Thus, Klein's trick is also a trick to obtain an invariant of odd degree from invariants of even degree.

Part of the beauty of Klein's generators lies in the fact that they have interesting geometric interpretations. The curve $f = 0$ in $\mathbb{P}^2(\mathbb{C})$ is, as noted, the modular curve of level 7 embedded by a natural basis for its holomorphic 1-forms. The curve $\nabla = 0$ is the locus of all points in the plane whose polar conics with respect to the quartic $f = 0$ are singular. The curve $C = 0$ is the locus of all points in the plane whose polar lines with respect to the Hessian $\nabla = 0$ are tangent to their polar conics with respect to the Klein curve $f = 0$. Furthermore, the curve $C = 0$ meets the Klein curve at the points of contact of its 28 bitangents. The curve $K = 0$ is the locus of all points whose polar lines with respect to $f = 0$, $\nabla = 0$ and $C = 0$ are concurrent. It also may be described in the following way: the group $\mathrm{PSL}_2(\mathbb{F}_7)$ has 21 elements of order 2. Each such involution fixes a projective line in $\mathbb{P}^2(\mathbb{C})$ as well as a point. Thus, the 21 lines associated to the 21 involutions form a reducible curve of degree 21 invariant under the group. Since there is only one invariant curve of degree 21, it must be $K = 0$. In particular, K is the product of 21 linear factors.

Klein also knew how to write the quartic f as a 4×4 determinant whose entries are linear forms in x, y, z. This fact may be expressed by saying that one may associate to f a net of quadrics in projective space and the Klein curve is the locus of singular quadrics. The locus of the singular points of the singular quadrics is a twisted curve of degree 6 and genus 3 isomorphic to the Klein curve.

Henceforth, we will freely use the notation introduced in the Appendix (Section 8). The reader is strongly advised to read the leisurely discussion there before proceeding, if only to gain passive knowledge of the relevant notation. However, to the more adventurous readers who prefer jungles to sidewalks, we offer the list of notation below as a machete. To facilitate such an index, groups of paragraphs of Section 8 have been numbered. Some notation is listed more than once, signifying that it has been redefined, specialized or generalized. This is especially the case for the Weil representation which is defined according to [Weil 1964] in 8.9 and denoted r_Γ, adapted to the case of finite symplectic groups $\mathrm{Sp}_{2n}(\mathbb{F}_p)$ in 8.12 and denoted r', composed with automorphisms σ_ν for $\nu\in\mathbb{F}_p^\times$ in 8.16 and denoted r'_ν, allowed to act on tensor powers of the version of 8.16 in 8.18 without change of notation, restricted to the subspaces $V_\nu^+=V^+$, $V_\nu^-=V^-$ of even and odd functions in 8.21 and denoted $\rho_\nu^\pm$, restricted to the symplectic groups $\mathrm{Sp}_{2n}(\mathbb{F}_q)$ of odd characteristic in 8.25 and denoted r'_ν with ν still a nonzero element of $\mathbb{F}_p$, and finally generalized to the case where ν is a nonzero element of $\mathbb{F}_q$ in 8.28. Derived notation such as ρ_ν^ε is not explicitly redefined in each context and the reader is expected to be able to make the necessary modifications without difficulty.

8.1	$G,\ G^*,\ \mathbb{T},\ A(G),\ \langle\cdot,\cdot\rangle$
8.2	$\mathbb{T}_0,\ A_0(G)$
8.3.1	$t_0(f)$
8.3.2	$d_0(\alpha)$
8.3.3	$d_0'(\gamma)$
8.4	$B(G)$
8.5	$L_2(G),\ U$
8.6	$\mathbf{A}(G),\ \mathbf{B}_0(G),\ \pi$
8.7.1	$\boldsymbol{t}_0(f)$
8.7.2	$\boldsymbol{d}_0(\alpha),\ \lvert\alpha\rvert$
8.7.3	$\boldsymbol{d}_0'(\gamma),\ \Phi^*,\ \lvert\gamma\rvert$
8.9	$B_0(G,\ \Gamma),\ r_\Gamma$
8.10	$\mathrm{Sp}(G),\ \mathrm{Sp}'(G)$
8.11	$\mathbf{B}_1(G),\ (E),\ V^+,\ V^-,\ S^+,\ S^-$
8.12	$G,\ \chi,\ \phi,\ Sp''(G),\ r'$
8.14	$G,\ \left(\frac{\cdot}{p}\right)$
8.15	$\nu,\ \sigma_\nu,\ s_\nu,\ s$
8.16	r'_ν
8.18	$r',\ r'_\nu,\ \boldsymbol{t}_0(f),\ \boldsymbol{d}_0(\alpha),\ \boldsymbol{d}_0'(\gamma)$
8.19	$\mathcal{T}$
8.21	$\rho_\nu^+,\ \rho_\nu^-,\ V_\nu^+,\ V_\nu^-$
8.24	$\mathcal{T}_{a,\ b}$
8.25	$r',\ r'_\nu$
8.26	$A^\#(G),\ [\cdot,\cdot],\ \mathrm{tr},\ \tau$
8.28	$\sigma_\nu,\ s_\nu,\ r'_\nu$
8.31	$\tau_a,\ \langle\cdot,\cdot\rangle,\ \mathcal{Q},\ \mathcal{Q}^+,\ \mathcal{Q}^-$

4. Conjectural Generators of the Bicycle of Invariants of $\mathrm{Aut}(\mathbb{F}_q) \cdot \mathrm{SL}_2(\mathbb{F}_q)$

In this section, we will try to provide a general context for the various tricks just studied. Let $q = p^r$ be an odd prime power and let ν be a nonzero element of $\mathbb{F}_q$. In [Adler 1992a; 1994] I showed (cf. 8.27–8.28) that there is a unique (up to scalar multiple) 3-tensor on $L_2(\mathbb{F}_q)$ invariant under the Weil representation r'_ν, and I wrote it down explicitly in general. We denote this 3-tensor by Θ. Let ε be the quadratic character of -1 in the finite field $\mathbb{F}_q$ and let η be the quadratic character of -2 in $\mathbb{F}_q$. Then Θ actually arises from an invariant 3-tensor on V_ν^+ if $\varepsilon = 1$ and on V_ν^- if $\varepsilon = -1$. If we abuse notation by identifying ε with its sign, we can say that Θ arises from a ρ_ν^ε-invariant 3-tensor on V_ν^ε. Further, Θ is a symmetric 3-tensor if $\eta = 1$ and is an alternating 3-tensor if $\eta = -1$. We can express this by saying that Θ is *η-symmetric*.

Regarding Θ as a 3-tensor on V_ν^ε, one can ask for the group of linear transformations of V_ν^ε which preserve Θ. In [Adler 1994], it was shown that for $q \geq 11$ the automorphism group is generated by the group

$$\rho_\nu^\varepsilon(\mathrm{SL}_2(\mathbb{F}_q) \cdot \mathrm{Aut}(\mathbb{F}_q))$$

and the group of scalar multiplications by cube roots of unity, provided that $q \neq 13$. Here, $\mathrm{Aut}(\mathbb{F}_q)$ denotes the Galois group of $\mathbb{F}_q$ over $\mathbb{F}_p$ and $\mathrm{SL}_2(\mathbb{F}_q) \cdot \mathrm{Aut}(\mathbb{F}_q)$ denotes the semidirect product of $\mathrm{Aut}(\mathbb{F}_q)$ and $\mathrm{SL}_2(\mathbb{F}_q)$. If $q = 13$, then the automorphism group is the complex Lie group G_2.

We now have 4 cases, according to the value of q modulo 8; these cases will be denoted 1, 3, 5, 7.

Case 1: If q is congruent to 1 modulo 8, we have $\eta = \varepsilon = 1$. We can write q in the form

$$q = 8n + 1.$$

The dimension of V^+ is $4n + 1$ and Θ is a cubic form on V^+. On the other hand, since q is congruent to 1 modulo 4, the representation of $\mathrm{SL}_2(\mathbb{F}_q)$ on V_ν^+ is orthogonal. The invariant quadratic form $\mathcal{Q}^+$ is given explicitly at the end of the appendix. According to the Bicycle Conjecture, we expect Θ and $\mathcal{Q}^+$ to generate the bicycle of invariants.

We could also have used Klein's tricks to produce the other conjectural generators, but it is clearly better to use the invariant quadratic $\mathcal{Q}^+$ as long as it is handy. The same remark applies, *mutatis mutandis*, in the case $q = 8n + 5$ below.

Case 3: If q is congruent to 3 modulo 8, then we have $\eta = 1$, $\varepsilon = -1$ and we can write q in the form

$$q = 8n + 3.$$

The dimension of V^- is $4n + 1$ and Θ is a cubic form on V^-. The Hessian of Θ is a form of degree $4n + 1$. If 3 doesn't divide $4n + 1$ then the Bicycle Conjecture implies that the bicycle of invariants is generated by Θ and its Hessian. If 3 does divide $4n+1$ then the bordered Hessian has degree

$$(4n-1)+2(4n)=12n-1,$$

which is not divisible by 3. In this case, the Bicycle Conjecture implies that the bicycle of invariants is generated by Θ and the bordered Hessian.

At the end of this section, we will describe a way of producing an invariant of degree $(q+1)/4$ from Θ by using certain intertwining operators.

REMARK 4.1. We have tacitly assumed that neither the Hessian determinant nor the bordered Hessian determinant is zero. We will make that assumption without explicit mention in all that follows. However, it certainly needs to be checked in any test of these conjectures. In cases 5 and 7 below, we will further assume that the invariants Ψ and Ψ' have automorphism group no bigger than that of Θ. In the case where q is a prime p of the form $4m+3$, this follows from the fact [Adler 1994] that $\mathrm{PSL}_2(\mathbb{F}_p)$ is a maximal algebraic subgroup of $\mathrm{PSL}_n(\mathbb{C})$ for $n=(p\pm 1)/2$. If p is congruent to 1 modulo 4 and $p\neq 13$, then the only algebraic subgroups of $\mathrm{PSL}_n(\mathbb{C})$ that could contain $\mathrm{PSL}_2(\mathbb{F}_p)$ are the orthogonal group $O(\frac{p+1}{2},\mathbb{C})$, if $\varepsilon=1$, or the symplectic group $\mathrm{Sp}(\frac{p-1}{2},\mathbb{C})$, if $\varepsilon=-1$. Of these two possibilties, only the orthogonal group has any polynomial invariants. Any homogeneous invariant for the orthogonal group is a power of the invariant quadratic form and in particular has even degree. Since the degree of Ψ is $4n+2$, it is conceivable that it is a power of the quadratic form. So we are assuming that this is not the case. However, as long as one is conjecturing, one might as well conjecture that the automorphism group is well behaved even when q is not a prime.

In the two preceding cases, we had $\eta=1$, which meant that the 3-tensor Θ was a cubic polynomial. In the remaining cases $\eta=-1$, which implies that the 3-tensor is alternating. We must therefore rely on different methods to produce invariants. However, we can still use Θ for that purpose by means of various tricks. By combining these tricks with those of Felix Klein, we can handle the remaining cases without difficulty.

Case 5: If q is congruent to 5 modulo 8, then we have $\eta=-1$ and $\varepsilon=1$ and we can write

$$q=8n+5.$$

In this case, Θ is an alternating 3-tensor on V^+. Since we will now deal with various Weil representations and we will want to keep track of them, we will say instead that it is an alternating 3-tensor on V^+_ν. We can then regard Θ as an equivariant mapping from V^+_ν to $\bigwedge^2 V^+_{-\nu}$. Since the dimension $4n+3$ of $V^+_{-\nu}$ is odd, we cannot obtain a nonzero invariant by composing with the Pfaffian. However, we can instead use the fundamental intertwining operator $\mathcal{T}$ to map $\bigwedge^2 V^+_{-\nu}$ isomorphically onto $\mathrm{Sym}^2 V^-_{-2\nu}$. Composing $\mathcal{T}\circ\Theta$ with the determinant on $\mathrm{Sym}^2 V^-_{-2\nu}$, we obtain an invariant Ψ of degree $4n+2$ on $V^- + \nu$. Since q is congruent to 1 modulo 4, we know that there is also an invariant $\mathcal{Q}^+$ of degree 2. If we assume that Ψ is nonzero and has automorphism group equal to $r'(\mathrm{SL}_2(\mathbb{F}_q) \cdot \mathrm{Aut}(\mathbb{F}_q))$ modulo scalars, then according to the Bicycle Conjecture, we can generate every invariant of even degree on V^+_ν from Ψ and $\mathcal{Q}^+$ using bicycle operations. However, there are certainly invariants of odd degree and we would like to get them too. It is enough, assuming the Bicycle Conjecture, to get just one of them. We can do that by adopting the trick used by Felix Klein to get his invariant K of degree 21 for $\mathrm{PSL}_2(\mathbb{F}_7)$, that is, by taking a Jacobian determinant. Indeed, once we have used bicycle operations to generate $4n+3$ algebraically independent forms of even degree starting with Ψ and $\mathcal{Q}^+$, we can then take their Jacobian determinant to get a form of odd degree. Using it and bicycle operations, we get the full bicycle of invariants.

We remark that the case $q = 13$ requires some additional concern since the automorphism group of Θ on V^+ is the complex Lie group G_2 in that case. However, since the construction of the proposed bicycle generators involves the use of the intertwining operator $\mathcal{T}$, which is only invariant under the smaller group $\mathrm{SL}_2(\mathbb{F}_{13})$, we don't have to worry about G_2. A similar phenomenon occurs in connection with the case $q = 7$, which will be discussed below.

The remaining case where q is congruent to 7 modulo 8 requires more care but requires no more than the 3-tensor Θ, the more general intertwining operators $\mathcal{T}_{a,b}$ and the tricks of Felix Klein. We begin by using $\mathcal{T}_{a,b}$ to produce a "twisted" version of the 3-tensor Θ. After that, we will turn to the details of Case 7.

By tensoring $\mathcal{T}_{a,b}$ with the identity operator on $L_2(G)$, we obtain an intertwining operator between

$$r'_{\nu,\nu,\nu} = r'_\nu \otimes r'_\nu \otimes r'_\nu \quad \text{and} \quad r'_{\mu,\mu,\nu} = r'_\mu \otimes r'_\mu \otimes r'_\nu,$$

where $\mu = (a^2 + b^2)\nu$. We will denote this intertwining operator by $\mathcal{T}_{a,b} \otimes 1$. Using the intertwining operator $\mathcal{T}_{a,b} \otimes 1$, we can regard the invariant 3-tensor Θ as an invariant 3-tensor, denoted Θ', for the representation $r'_{\mu,\mu,\nu}$ or, what is the same, as an equivariant mapping from r'_ν to $r'_{-\mu,-\mu}$. The existence of this invariant 3-tensor can also be shown using the same proof given in [Adler 1992a, Theorem 1] for the existence of Θ; even the computation is the same, up to the order of the terms to be added. It is also possible to write the invariant 3-tensor Θ' down explicitly by suitably adapting the methods and results of [Adler 1992a]. As in that paper, we can write Θ' in the form

$$\sum \kappa(x,y,z)\delta_x \otimes \delta_y \otimes \delta_z,$$

where δ_t denotes the delta function at t for all $t \in \mathbb{F}_q$, where $\kappa(x,y,z)$ is a complex number and where the summation runs over all elements (x,y,z) of $\mathbb{F}_q^3$ By acting on Θ' with the element $r'_{\mu,\mu,\nu}$, we see that $\kappa(\mu,\mu,\nu)$ vanishes unless

$$\mu(z^2 + y^2) + \nu z^2 = 0.$$

Let Q denote the space of binary quadratic forms with entries in $\mathbb{F}_q$, which we identify with their matrices. Let i, j be elements of $\mathbb{F}_q$ such that $i^2 + j^2 \neq 0$. Define the bijective linear mapping

$$\lambda : \mathbb{F}_q^3 \to Q$$

by

$$\lambda(x, y, z) = \begin{pmatrix} \nu z + ix + jy & jx - iy \\ jx - iy & \nu z - ix - jy \end{pmatrix}.$$

The determinant of $\lambda(x, y, z)$ is easily seen to be

$$\nu^2 - (i^2 + j^2)(x^2 + y^2).$$

Therefore, if we choose i, j such that $i^2 + j^2 = -\nu\mu$, we see that the coefficient $\kappa(x, y, z)$ vanishes unless the determinant of $\lambda(x, y, z)$ is zero. It follows as in [Adler 1992a] that we can take the coefficient $\kappa(x, y, z)$ to be given by

$$\kappa(x, y, z) = \begin{cases} 0 & \text{if } \lambda(x, y, z) \text{ has rank } 2, \\ 0 & \text{if } \lambda(x, y, z) = 0, \\ 1 & \text{if } \lambda(x, y, z) \text{ is the square of a linear form,} \\ -1 & \text{otherwise.} \end{cases}$$

The quadratic form $\lambda(x, y, z)$ is given by

$$(\nu z + ix + jy)s^2 + 2(jx - iy)st + (\nu z - ix - jy)t^2.$$

We cannot expect the coefficients $\kappa(x, y, z)$ to have nice properties under all permutations of x, y, z since the quadratic form $\mu(x^2 + y^2) + \nu z^2$ doesn't. But it is reasonable to expect good behavior under interchange of x, y. Indeed, we have

$$\kappa(y, x, z) = \left(\frac{-2\mu\nu}{\mathbb{F}_q}\right) \kappa(x, y, z), \tag{4.1}$$

where the coefficient of $\kappa(x, y, z)$ on the right hand side is the quadratic character of $-2\mu\nu$ in the finite field $\mathbb{F}_q$. If we take μ to be a square and ν to be a nonsquare in $\mathbb{F}_q$ then the coefficient of $\kappa(x, y, z)$ is the negative of the quadratic character of -2 in $\mathbb{F}_q$, that is, the symmetry properties of this "twisted" 3-tensor invariant under switching x, y are the *opposite* of those of the original 3-tensor Θ. As for the behavior of $\kappa(x, y, z)$ under replacing one or more of x, y, z by their negatives, we find that

$$\kappa(-x, y, z) = \kappa(x, -y, z) = -\left(\frac{-1}{\mathbb{F}_q}\right) \kappa(x, y, z) \tag{4.2}$$

and

$$\kappa(x, y, -z) = \left(\frac{-1}{\mathbb{F}_q}\right) \kappa(x, y, z). \tag{4.3}$$

We may regard the $r'_{\mu,\mu,\nu}$-invariant 3-tensor Θ' as an equivariant mapping from r'_ν to $r'_{-\mu,-\mu}$. It now follows from (4.1), (4.2) and (4.3) that Θ' gives rise to an equivariant mapping

$$V_\nu^- \to \bigotimes\nolimits_\eta^2 V_{-\mu}^+$$

if $\varepsilon = -1$ and to an equivariant mapping

$$V_\nu^+ \to \bigotimes\nolimits_\eta^2 V_{-\mu}^-$$

if $\varepsilon = 1$, where $\bigotimes_\eta^2$ is defined by

$$\bigotimes\nolimits_\eta^2 = \begin{cases} \mathrm{Sym}^2 & \text{if } \eta = -1 \\ \bigwedge^2 & \text{if } \eta = 1 \end{cases}$$

Case 7: In the case at hand, we have q congruent to 7 modulo 8. Therefore we have $\eta = \varepsilon = -1$ and we can write q in the form

$$q = 8n - 1.$$

The twisted invariant 3-tensor then gives us an equivariant mapping

$$V_\nu^- \to \mathrm{Sym}^2 V_{-\mu}^+.$$

Composing this mapping with the determinant, we obtain an invariant Ψ' of degree $4n$ on V_ν^-. In the special case $p = 7$, the invariant Ψ' is none other than Klein's quartic

$$x^3y + y^3z + z^3x,$$

up to a scalar factor. It is therefore not surprising that Klein's tricks work in this case as well to give us conjectural generators of the bicycle of invariants. Indeed, the Hessian of Ψ' has degree

$$(4n - 2)(4n - 1)$$

and the greatest common divisor of the degrees of Ψ' and of its Hessian is 2. Therefore, according to the Bicycle Conjecture, the bicycle of all invariants of even degree of $\mathrm{SL}_2(\mathbb{F}_q)$ in V_ν^- is generated by Ψ' and its Hessian. Since the space V_ν^- has odd dimension, we can also get an invariant of odd degree by using Klein's trick of taking the Jacobian determinant of $4n - 1$ algebraically independent invariants of even degree.

We note here that in the case $q = 7$, the automorphism group of Θ on V^- is $\mathrm{SL}_3(\mathbb{C})$, not $\mathrm{PSL}_2(\mathbb{F}_7)$. However, since the construction of Klein's quartic from Θ involves the use of the intertwining operator $\mathcal{T}_{a,b}$, which is not invariant under $\mathrm{SL}_3(\mathbb{C})$, we are cut down to the smaller group $\mathrm{PSL}_2(\mathbb{F}_7)$. This is similar to what happened in the case $q = 13$ with the group G_2.

THE Θ CONJECTURE 4.2. *The following table gives conjectural bicycle generators for* $\mathrm{SL}_2(\mathbb{F}_q)$ *acting on* V_ν^ε:

$q\equiv 1 \pmod 8$	Θ, $\mathfrak{Q}^+$
$q\equiv 3 \pmod 8$, $\not\equiv 1 \pmod 6$	Θ, Hessian(Θ)
$q\equiv 3 \pmod 8$, $\equiv 1 \pmod 6$	Θ, Bordered Hessian(Θ)
$q\equiv 5 \pmod 8$	Ψ, $\mathfrak{Q}^+$, Jacobian
$q\equiv 7 \pmod 8$	Ψ', Hessian(Ψ), Jacobian

Thus we have conjectural generators of the bicycle of invariants of $\mathrm{SL}_2(\mathbb{F}_q)$ in V^ε in every case. Unfortunately, our methods so far tell us essentially nothing about the case of the other component of the Weil representation. In the next section, we will try to improve the situation a little. We will merely close this section with two simple remarks.

REMARK 4.3. Our examination of the invariant 3-tensor Θ and the invariant "twisted" 3-tensor Θ' provides us with an essentially unique nonassociative algebra structure on $L_2(\mathbb{F}_q)$ invariant under $\mathrm{SL}_2(\mathbb{F}_q)\cdot\mathrm{Aut}(\mathbb{F}_q)$.

REMARK 4.4. Our construction of Klein's quartic from Θ shows, among other things, how to write Klein's quartic explicitly as a symmetric 4×4 determinant whose entries are linear forms on V_ν^-. Klein [1879a; 1890–92, vol. II, ch. V] gave an explicit representation of his own quartic and studied the geometry of the associated curve in projective space as well as the plane curve. This study was taken further by H. F. Baker [1935] and especially by W. L. Edge [1947], who made a detailed study of the geometry of the net of quadrics determined by Klein's determinantal representation. In view of the fact that this determinantal representation is herein generalized to the case $q=8n-1$, it appears that [Edge 1947] might be a source of considerable inspiration for what to prove in the general case. We also note that Klein's cubic form

$$v^2w+w^2x+x^2y+y^2z+z^2v,$$

which arises in our setting as the invariant 3-tensor Θ in the case $q=11$, can be expressed as the Pfaffian of an alternating 6×6 matrix whose entries are linear forms in 5 variables. Indeed, in the general case $q=8n+3$, we can regard the invariant 3-tensor Θ as an equivariant mapping from V_ν^- to $\mathrm{Sym}^2 V_{-\nu}^-$. Composing this equivariant mapping with the fundamental intertwining operator $\mathfrak{T}$, we obtain an equivariant mapping from V_ν^- to $\bigwedge^2 V_{-2\nu}^+$. Composing this mapping with the Pfaffian, we obtain an invariant of degree $(q+1)/4$ on V_ν^-. In case $q=11$, that degree is 3 and we have expressed our unique cubic invariant as a Pfaffian, as promised. This Pfaffian representation therefore appears to be on an equal footing, from our general point of view, with the determinantal representation of Klein's quartic. Explicitly, Klein's cubic is the Pfaffian of the following skew-symmetric matrix:

$$\begin{pmatrix} 0 & v & w & x & y & z \\ -v & 0 & 0 & z & -x & 0 \\ -w & 0 & 0 & 0 & v & -y \\ -x & -z & 0 & 0 & 0 & w \\ -y & x & -v & 0 & 0 & 0 \\ -z & 0 & y & -w & 0 & 0 \end{pmatrix}.$$

Finally, we note that in the special case $q = 11$, every point of Klein's cubic threefold in projective 4 space $\mathbb{P}(V_\nu^-)$ determines a projective line in projective 5 space $\mathbb{P}(V_\nu^+)$ and the locus swept out by these lines is defined by the unique quartic invariant of $\mathrm{SL}_2(\mathbb{F}_{11})$ in V_ν^+. As noted in [Adler 1997], the singular locus of that quartic is birationally equivalent to the modular curve $X(11)$.

5. Conjectural Generators of the Bicycle of Invariants of $\mathrm{Sp}(\mathbb{F}_p^r)$ on V_ν^+

As in the preceding section, let p be an odd prime number and let r be a positive integer. If $p = 3$, we will assume that $r > 1$. We have the Weil representation r'_ν of $\mathrm{Sp}(\mathbb{F}_p^r)$ on $L_2(\mathbb{F}_p^r)$ and its tensor powers, which are also denoted r'_ν. According to Lemma 8.23, the canonical intertwining operator $\mathcal{T}$ maps $\mathrm{Sym}^2(V_\nu^+)$ onto $\mathrm{Sym}^2(V_{2\nu}^+)$. Since the dual space of $\mathrm{Sym}^2(V_\nu^+)$ is $\mathrm{Sym}^2(V_{-\nu}^+)$, the canonical interwining operator can be viewed as mapping $\mathrm{Sym}^2(V_\nu^+)$ onto its dual if -2 is a square modulo p. We will assume that this is the case. More precisely, we can use the intertwining operator $\mathcal{T}_{a,a}$ where a is an element of $\mathbb{F}_p$ such that $2a^2 = -1$ to map $\mathrm{Sym}^2(V_\nu^+)$ onto $\mathrm{Sym}^2(V_{-\nu}^+)$. We therefore obtain a linear form on $\mathrm{Sym}^2(V_\nu^+) \otimes \mathrm{Sym}^2(V_\nu^+)$ whose restriction to $\mathrm{Sym}^4(V_\nu^+)$ is a $\mathrm{Sp}(\mathbb{F}_p^r)$-invariant quartic form Ω on the even part of the Weil representation.

Theorem 5.1. *The quartic form Ω is nonzero.*

Proof. As in the proof of Lemma 8.23, we may regard an element of $\mathrm{Sym}^2(V_\nu^+)$ as a function $f(x, y)$ such that

$$f(-x, y) = f(x, -y) = f(y, x)$$

for all $x, y \in \mathbb{F}_p^r$. We want to show that for some $f \in \mathrm{Sym}^2(V_\nu^+)$ we have

$$\Omega(f) \neq 0$$

or, what is the same, that

$$(\mathcal{T}_{a,a} f)(f) \neq 0.$$

The linear form

$$\beta : \mathrm{Sym}^2(V_\nu^+) \otimes \mathrm{Sym}^2(V_{-\nu}^+) \to \mathbb{C}$$

given by

$$\beta(f, g) = \sum f(x, y) g(x, y),$$

where the summation runs over all $x,y\in\mathbb{F}_p^r$, is a nondegenerate pairing which is invariant under the action of

$$\mathrm{Sym}^2(\rho_\nu^+)\otimes\mathrm{Sym}^2(\rho_{-\nu}^+).$$

Therefore we only have to check that for some f in $\mathrm{Sym}^2(V_\nu^+)$ of the form $g\otimes g$ with $g\in V_\nu^+$ we have

$$\sum f(x,y)f(ax+ay,ax-ay)\neq 0,$$

where the summation runs over all $x,y\in\mathbb{F}_p$. We will take f to be the function which is 1 at $(0,0)$ and 0 elsewhere. We then have

$$\sum f(x,y)f(ax+ay,ax-ay)=1,$$

which proves that the quartic invariant is nonzero. □

We can write the quartic invariant explicitly as follows. Let $Y\in V_\nu^+$. Then the invariant is

$$\sum(Y\otimes Y)(x,y)(Y\otimes Y)(ax+ay,\ ax-ay)=\sum Y(x)Y(y)Y(ax+ay)Y(ax-ay).$$

For example, the unique quartic invariant of $\mathrm{SL}_2(\mathbb{F}_7)$ in 4 variables and of $\mathrm{SL}_2(\mathbb{F}_{11})$ in 6 variables arise in this way. For the case $p=3$ and $r=2$, the quartic invariant was discovered by Burckhardt [1893] and studied in detail by various authors, such as Baker [1935] and Coble [1917]. In [Adler and Ramanan 1996] we generalized Burckhardt's quartic to the case $p=3$ and $r>1$ and proved that it was the unique quartic invariant for this representation. In this case, the quartic can be written in the following way. For each element u of $\mathbb{F}_3^r$, introduce a variable Y_u with the provision that $Y_{-u}=Y_u$. It is the same to introduce the variable Y_0 and, for each one-dimensional $\mathbb{F}_3$ subspace λ of $\mathbb{F}_3^r$, a variable Y_λ. Then the invariant is given by

$$\sum_{u,v\in\mathbb{F}_3^r}Y_uY_vY_{u+v}Y_{u-v},$$

which can also be written as

$$Y_0^4+8Y_0\sum Y_\lambda^3+48\sum_{\pi}\prod_{\lambda\subset\pi}Y_\lambda,$$

where the first summation runs over all one-dimensional subspaces of the $\mathbb{F}_3$ vector space $\mathbb{F}_3^r$, the second summation runs over all two-dimensional subspaces π and where the product runs over all one-dimensional subspaces λ contained in a given two-dimensional subspace π.

The Hessian of Ω will be denoted Υ. The form Υ has degree p^r+1. If the Bicycle Conjecture could be applied in this case, we would obtain obtain the following conjecture:

The Ω Conjecture 5.2. *If p be an odd prime and let r be a positive integer. If $p = 3$ assume that $r > 1$. Let m be the greatest common divisor of 4 and $p^r + 1$. If p^r is congruent to 1 modulo 8 then let $M = 2$. Otherwise let $M = 4$. Then every invariant of ρ_ν^+ of degree divisible by M is obtained from Ω and Υ by bicycle operations.*

In case $(p^r + 1)/2$ is odd we can get an invariant of odd degree from a Jacobian determinant and use it to produce all invariants.

Implicit in the above conjectures is the assumption that Υ is not zero. Assuming this to be the case, one also must be certain that the forms Ω and Υ have no automorphisms in common other than those of $\mathrm{Sp}(\mathbb{F}_p^r)$, modulo scalars. That is verified in the following lemma.

Lemma 5.3. *The automorphism group of Ω is generated by $\rho_\nu^+(\mathrm{Sp}(\mathbb{F}_p^r))$ and by scalar multiplication by fourth roots of unity. Modulo scalars the group is precisely* $\mathrm{PSp}(\mathbb{F}_p^r)$.

Proof. In [Adler 1994], it is shown that the group $\mathrm{PSp}(\mathbb{F}_p^r)$ is a maximal algebraic subgroup of the group of collineations of $\mathbb{P}^N(\mathbb{C})$, where $N = (q^r \pm 1)/2$. Therefore, since Ω has degree 4, every automorphism of Ω is the product of an element of $\mathrm{Sp}(\mathbb{F}_p^r)$ and scalar multiplication by a fourth root of unity. $\square$

The Ω Conjecture give us conjectural generators of the bicycle of invariants of even degree for $\mathrm{Sp}(\mathbb{F}_p^r)$ on V_ν^+. If p^r is congruent to 3 modulo 4 then the center of $\mathrm{Sp}(\mathbb{F}_p^r)$ acts as -1 on V_ν^+ and all invariants are necessarily of even degree. So if p^r is congruent to 3 modulo 4, these two conjectures give us conjectural generators for the full ring of invariants of ρ_ν^+.

In the special case where $r = 1$, we are dealing with the group $\mathrm{SL}_2(\mathbb{F}_p)$. If p is congruent to 3 modulo 4, the quadratic character of -1 modulo p is $\varepsilon = -1$. In this case, the methods of the preceding section give us conjectural generators of the ring of invariants (or those of even degree, at least) in V_ν^-. The methods of this section give us conjectural generators of the ring of invariants of V_ν^+ as well. Thus we have made some progress towards completing our list of conjectures. In this section, we have also opened the door to the invariants of finite symplectic groups in general in the Weil representation. It is desirable to extend the conjectures to these cases as well.

We close this section by noting some typographical and other errors in [Adler 1994]. In the statement of Lemma 7.1 on p. 2354, the group denoted $\mathrm{PSp}_m(R)$ in (1) and (1′) should be denoted $\mathrm{PSp}_{2m}(R)$. Similarly, in the statement of Theorem 7.2 on p. 2355, the group denoted $\mathrm{PSp}_m(R)\cdot A$ in (1) and (1′) should be denoted $\mathrm{PSp}_{2m}(R)\cdot A$. On the same page, in the proof of Theorem 7.2, the groups denoted $\mathrm{PSp}_m(R)$, $\mathrm{PSp}_r(\mathbb{F}_p)$, $\mathrm{PSp}_m(\mathbb{F}_R)$ and $\mathrm{Sp}_m(\mathbb{F}_R)$ should be respectively be denoted $\mathrm{PSp}_{2m}(R)$, $\mathrm{PSp}_{2r}(\mathbb{F}_p)$, $\mathrm{PSp}_{2m}(\mathbb{F}_R)$ and $\mathrm{Sp}_{2m}(\mathbb{F}_R)$. Also, the statement of Theorem 8.2 on p. 2360 and the paragraph preceding it should read:

If η is ± 1, we will say that a tensor T is **η-symmetric** if $\eta = 1$ and T is symmetric or if $\eta = -1$ and T is skew-symmetric. If $\varepsilon = \pm 1$, then by the **ε-part** of the Weil representation, we will mean the even part if $\varepsilon = 1$ and the odd part if $\varepsilon = -1$.

Theorem (8.2): *Suppose $q \geq 11$. Let ε equal the quadratic character of -1 in $\mathbb{F}_q$ and let η equal the quadratic character of -2 in $\mathbb{F}_q$. Let $n = (q+\varepsilon)/2$. Denote by Θ the unique η-symmetric 3-tensor on the ε part of the Weil representation of $\mathrm{SL}_2(\mathbb{F}_q)$. Then the group of collineations which preserve the 3-tensor is isomorphic to $\mathrm{PSL}_2(\mathbb{F}_q)\cdot\mathrm{Aut}(\mathbb{F}_q)$ unless $q = 13$, in which case the group in question is $G_2(\mathbb{C})$.*

Finally, I wish to correct my comments about the work of van der Geer in [Adler 1994]. Since the appearance of that article, I have received a copy of a letter he has written in which he acknowledges my conversation with him about the quartic invariant and its explicit form. He explains that he had obtained the explicit form of the quartic invariant of $\mathrm{Sp}_{2n}(\mathbb{F}_3)$ independently at about the same time that Ramanan and I did. Accordingly, I would like to apologize for my remarks in [Adler 1994]. I hope that this apology will serve to correct the negative impressions that my comments may have caused about the character and accomplishments of a mathematician in whose work I have found so much to admire.

6. Geometric Constructions

According to Klein's Erlangen Program, geometry is the study of the properties of a set X which are preserved by the action of a group G on the set X. One important example is complex projective n-space with the group $\mathrm{PSL}_{n+1}(\mathbb{C})$ acting on it. We are familiar with this example, as it is quite standard. But suppose G is a finite group and ρ is a homomorphism from G into $\mathrm{SL}_{n+1}(\mathbb{C})$. Then G also acts on $\mathbb{P}^n(\mathbb{C})$ and gives rise to a different notion of geometry on the same set. It is quite instructive to try to articulate the difference between these two geometries.

In these examples, we can already see that the definition of geometry given above leaves certain important details unspecified. For example, in complex projective n-space, one can spend all one's time looking only at linear subspaces. Or, one can be an algebraic geometer and consider all algebraic loci in complex projective n-space. In either case, one has the same group acting but one is really considering two different kinds of geometry. Thus, we have left unspecified the kinds of objects one might want to focus on. In practice there will be various types of objects one studies in the geometry. For example, in the projective plane, one can study points, lines, triangles, conics and so forth.

Suppose G is a group acting on a set X and suppose that we have agreed on the types of objects we will consider in this geometry. If T is a type of object, we will denote by $T(X)$ the set of all objects of type T in X. Then G acts on the set $T(X)$. Suppose T_1 and T_2 are two types of object. By a *construction* of objects of type T_2 from objects of type T_1, we mean a G-equivariant mapping from $T_1(X)$ to $T_2(X)$.

Here we have to be careful, since in practice one has certain preferences as to what kind of mappings one will allow. For example, in the case of algebraic geometry in complex projective space, we would perhaps only allow polynomial mappings or rational mappings. So in our definition of a geometric construction, we really mean to assume that we are dealing with a certain category of mappings.

I would like to examine this notion of a construction more closely in the case of complex projective space. For definiteness and for simplicity, I want to focus on the question of constructing one hypersurface from another one. If d is a positive integer and V is a complex vector space, we will denote by $S_d(V)$ the vector space of forms of degree d on V. The set of hypersurfaces of degree d in the projective space $\mathbb{P}(V)$ of lines in V may then be roughly identified with the projective space $\mathbb{P}(S_d(V))$. I say roughly because a hypersurface does not uniquely determine the form which defines it, at least if we regard the matter set theoretically. For example, in $\mathbb{P}^2$, with homogeneous coordinates x, y, z, the forms x^3y and xy^3 define the same hypersurface but are not proportional. However, I am going to overlook this difficulty and pretend that the set of hypersurfaces of degree d in $\mathbb{P}^n$ is $\mathbb{P}(S_d(V))$. The difficulty disappears if one regards a hypersurface as a scheme instead of as a set, but I want to keep the discussion elementary.

Let d, e be positive integers and suppose that

$$F : \mathbb{P}(S_d(\mathbb{C}^{n+1})) \to \mathbb{P}(S_e(\mathbb{C}^{n+1}))$$

is a geometrical construction of hypersurfaces of degree e from hypersurfaces of degree d. Since we are doing algebraic geometry, that means we want the mapping F to be a rational mapping or a polynomial mapping. One problem with using rational mappings is that if one wishes to take the result of the construction F and apply another construction to it, say F', the result may not be defined. So geometric constructions don't really give us a category. On the other hand, by writing F out explicitly in terms of the coefficients of the general form of degree d and clearing denominators, we obtain a polynomial mapping

$$\tilde{F} : S_d(\mathbb{C}^{n+1}) \to S_e(\mathbb{C}^{n+1})$$

which lifts F. Since F is by definition equivariant for the action of $\mathrm{SL}_{n+1}(\mathbb{C})$, it follows that $\tilde{F}$ is a homogeneous mapping equivariant for the action of $\mathrm{SL}_{n+1}(\mathbb{C})$. In other words, $\tilde{F}$ is precisely what one classically called a *covariant*. More generally, if W is any representation space for $\mathrm{SL}_{n+1}(\mathbb{C})$, a homogeneous polynomial mapping from W to $S_e(\mathbb{C}^{n+1})$ equivariant for $\mathrm{SL}_{n+1}(\mathbb{C})$ would be called a covariant of degree e on W. The degree of the mapping is called the *order* of the covariant. In the special case where $e = 0$, a covariant is called an *invariant*.

We should also note that by using covariants, the difficulties of composing geometric constructions disappears: one can always compose polynomials.

Classically, one also considered loci as being defined by their families of tangent hyperplanes. Since a hyperplane is a point of the dual projective space, this approach amounts to a study of loci in the dual projective space. Asking for geometric constructions of these loci amounts to asking for equivariant mappings

$$S_d(\mathbb{C}^{n+1}) \to S_e(\mathbb{C}^{n+1^*}).$$

Such a mapping is called a *contravariant.* Finally, one classically considered relations between projective space and its dual which depend geometrically on a given hypersurface. This leads one to study equivariant mappings

$$S_d(\mathbb{C}^{n+1}) \to S_e(\mathbb{C}^{n+1}) \otimes S_f(\mathbb{C}^{n+1^*}),$$

which are called *mixed concomitants.*

Now that we understand a little better what we mean by a construction in classical projective algebraic geometry, let us examine the geometry imposed on $\mathbb{P}^n$ by a representation

$$\rho : G \to \mathrm{SL}_{n+1}(\mathbb{C})$$

of a finite group G. People who studied this kind of geometry were concerned not with all loci but only with loci invariant under the action of G. The reason this was not done in the case of classical projective geometry is that the group $\mathrm{SL}_{n+1}(\mathbb{C})$ acts transitively on $\mathbb{P}^n$ and there are no invariant loci. But with the finite group G, such loci exist in abundance and geometers have long delighted in studying them.

Suppose T_1, T_2 are types of objects in this geometry. A geometric construction of objects of type T_2 from objects of type T_1 is then a G-equivariant mapping

$$T_1(\mathbb{P}^n) \to T_2(\mathbb{P}^n).$$

As before, we need to specify the category of mappings we are using and again we will side with the algebraic geometers in choosing rational or polynomial maps. But more important is the following observation: since we are only interested in invariant loci, the group G acts trivially on $T_1(\mathbb{P}^n)$ and $T_2(\mathbb{P}^n)$. Therefore *every* mapping is equivariant. The notion of a geometric construction apparently loses all of its content. To put the matter bluntly, it is as easy to do geometric constructions in this geometry as it is to write poetry in Pig Latin. (Good poetry is, of course, another matter.)

While this conclusion is at first rather disconcerting, we may take heart in the observation that herein lies one of the ways we can articulate the difference between the geometry imposed by G and classical projective algebraic geometry. Indeed, we may ask: when can a geometric construction in the G-geometry be effected by means of a construction in classical projective algebraic geometry?

For example, recall Klein's generators of the ring of invariants of $\mathrm{PSL}_2(\mathbb{F}_7)$ in a three-dimensional complex representation, which we discussed in Section 3. Klein started with the invariant quartic

$$f = x^3y + y^3z + z^3x$$

and then wrote down 3 other invariants ∇, C, K of degrees 6, 14, 21 explicitly. According to the geometry imposed on $\mathbb{P}^2$ by $\mathrm{PSL}_2(\mathbb{F}_7)$, the mere juxtaposition of f and C, for example, amounts to a geometric construction of C from f. We explore the difference between $\mathrm{PSL}_2(\mathbb{F}_7)$ geometry and classical plane projective geometry when we ask whether there is a covariant

$$S_4(\mathbb{C}^3) \to S_{14}(\mathbb{C}^3)$$

mapping f to C. And in fact, there is: Klein himself gave it when he expressed C as a constant times the 4×4 matrix obtained by bordering the matrix of second partials of f with the first partials of the Hessian ∇ of f.

Thus, one way of exploring the difference between these two geometries is to ask whether every invariant of $\mathrm{PSL}_2(\mathbb{F}_7)$ arises by applying a covariant to f. Since all of Klein's generators are given explicitly by covariants, it appears that the answer to this question is affirmative.

For another example, consider the cubic form

$$f_3 = v^2w + w^2x + x^2y + y^2z + z^2v,$$

also discovered by Klein [1879b]. It is the unique (up to constant multiple) cubic invariant of a five-dimensional irreducible complex representation of $\mathrm{SL}_2(\mathbb{F}_{11})$. I computed the generators and relations of the ring of invariants of this representation and found that it is generated by 10 polynomials

$$f_3, f_5, f_6, f_7, f_8, f_9, f_{10}, f_{11}, f_{12}, f_{14},$$

where f_n has degree n. I was able to express all of the invariants explicitly using covariants of f_3 except for f_{11}. For years, I didn't know whether it was expressible by covariants or not. But as we will see in Corollary 7.6 below, it is in fact expressible in this way, as are all of the invariants.

If we write down the matrix of second partial derivatives of Klein's cubic we find that up to a trivial factor of 2, it is

$$\begin{pmatrix} w & v & 0 & 0 & z \\ v & x & w & 0 & 0 \\ 0 & w & y & x & 0 \\ 0 & 0 & x & z & y \\ z & 0 & 0 & y & v \end{pmatrix}.$$

Its determinant is the invariant I have denoted f_5 and is the Hessian of f_3 up to a factor of 32.

Now consider the locus of all point $[v, w, x, y, z]$ in $\mathbb{P}^4$ for which this matrix has rank equal to 3. Saying that the rank is at most 3 amounts to writing down the 4×4 minors of the matrix and setting them to 0. That gives us a lot of quartics which define an algebraic locus in $\mathbb{P}^4$. On the other hand, it is not difficult to show, as Klein did, that there are no points $[v, w, x, y, z]$ for which the rank is less than 3. Therefore the rank 3 locus is an algebraic locus.

Felix Klein discovered the remarkable theorem that this locus is isomorphic to the modular curve $X(11)$ of level 11. Let me call this Theorem K. He also expressed this result by saying that $X(11)$ is isomorphic to the singular locus of the hypersurface $f_5 = 0$, that is, that $X(11)$ is the singular locus of the Hessian of the cubic $f_3 = 0$. Let me call this Theorem K′. I would like to mention that these two theorems do not say exactly the same thing, although it is not hard to show (as Klein did) that they are really equivalent. Meanwhile let me merely note that from Theorem K′, it is immediately apparently how the group $\mathrm{PSL}_2(\mathbb{F}_{11})$ acts on the modular curve $X(11)$. For f_3 is an invariant of $\mathrm{PSL}_2(\mathbb{F}_{11})$, its Hessian is likewise an invariant and therefore the singular locus of the Hessian is invariant under the group.

However one states the theorem, I have always found this to be an inspiring result. One naturally wonders whether one can generalize it. This problem has occupied me for a number of years.

Actually, Klein himself found a beautiful generalization of his theorem. Let $p \geq 5$ be a prime number. Denote by $L^2(\mathbb{F}_p)$ the p-dimensional complex vector space of all (square-integrable) complex valued functions on $\mathbb{F}_p$ with respect to counting measure, that is, all functions from $\mathbb{F}_p$ to the complex numbers. We can decompose $L^2(\mathbb{F}_p)$ as the direct sum of the space V^+ of even functions and the space V^- of odd functions. The space V^- has dimension $(p-1)/2$ and its associated projective space $\mathbb{P}(V^-)$ has dimension $(p-3)/2$. If f is a nonzero element of V^-, we will denote by $[f]$ the corresponding element of $\mathbb{P}(V^-)$, in keeping with the classical notation for homogeneous coordinates.

Klein discovered the following general result:

THEOREM 6.1. *The modular curve $X(p)$ is isomorphic to the locus of all $[f]$ in $\mathbb{P}(V^-)$ which for all w, x, y, z in $\mathbb{F}_p$ satisfy the identities*

$$\begin{aligned} 0 = {} & f(w+x)f(w-x)f(y+z)f(y-z) \\ & + f(w+y)f(w-y)f(z+x)f(z-x) \\ & + f(w+z)f(w-z)f(x+y)f(x-y). \end{aligned}$$

Thus, $X(p)$ is defined by a collection of quartics which we can write down explicitly. In the special case $p = 11$, we recover Klein's theorem about $X(11)$. In the case $p = 7$, we obtain the defining equation of the Klein curve,

$$x^3y + y^3z + z^3x = 0$$

As much as we may admire this theorem, it is natural to feel somewhat daunted by it. For even though we know the equations, there are an awful lot of equations and it isn't clear that they really do us any good. To persuade you otherwise, let me mention that in [Adler and Ramanan 1996, § 19] we looked closely at these equations and found that they have a simple geometric interpretation: they say that the modular curve $X(p)$ is the intersection of a Grassmannian and a 2-uply embedded projective space!

More precisely, consider the Weil representation of $\mathrm{SL}_2(\mathbb{F}_p)$ on $L_2(\mathbb{F}_p)$. Tensor this representation with itself and identify $L_2(\mathbb{F}_p) \otimes L_2(\mathbb{F}_p)$ with $L_2(\mathbb{F}_p^2)$. Define the operator T from $L_2(\mathbb{F}_p^2)$ to itself by

$$(T\Phi)(x, y) = \Phi\Big(\frac{x+y}{2}, \frac{x-y}{2}\Big).$$

Then one can show that T normalizes $\mathrm{SL}_2(\mathbb{F}_p)$ as a group of operators on $L_2(\mathbb{F}_p^2)$ and maps $\bigwedge^2(V^+)$ isomorphically onto $\mathrm{Sym}^2(V^-)$. Passing to projective spaces, we can use T to identify $\mathbb{P}(\bigwedge^2(V^+))$ with $\mathbb{P}(\mathrm{Sym}^2(V^-))$. Now, in $\mathbb{P}(\bigwedge^2(V^+))$ we have the Grassmannian Gr of complex 2-planes in V^+ and in $\mathbb{P}(\mathrm{Sym}^2(V^-))$ we have the image Ver of $\mathbb{P}(V^-)$ under the 2-uple embedding. Klein's equations say precisely that $X(p)$ is the intersection of Gr and Ver.

Incidentally, one immediate consequence of this interpretation is the otherwise non-obvious result that the modular curve $X(p)$ has a canonical $\mathrm{SL}_2(\mathbb{F}_p)$ invariant rank 2 vector bundle that it gets from the Grassmannian Gr. This vector bundle is considered in more detail in [Adler and Ramanan 1996, § 24].

If Klein already generalized his theorem about $X(11)$ (that is, Theorem K) to all p, why am I not satisfied? Well, look again at Theorem K′. It says that the modular curve $X(11)$ is the singular locus of the Hessian of $f_3 = 0$. In particular, it says that we can construct the modular curve $X(11)$ from the cubic invariant f_3. Now there is nothing in Klein's general theorem on $X(p)$ about any cubic. It just gives a bunch of quartic equations that define $X(p)$. On the other hand, Ramanan and I proved that whenever $p > 3$ is a prime congruent to 3 modulo 8 (e.g. the prime $p = 11$), there is a unique cubic invariant for the representation of $\mathrm{SL}_2(\mathbb{F}_p)$ on V^-. At least for such p, we have an invariant cubic hypersurface in $\mathbb{P}(V^-)$ and we have the modular curve $X(p)$. So we have the right to ask: can we construct the modular curve $X(p)$ geometrically from the cubic hypersurface for all such p?

More generally, for any $p > 3$ there is a unique 3-tensor Θ on $L_2(\mathbb{F}_p)$ invariant under the Weil representation of $\mathrm{SL}_2(\mathbb{F}_p)$, as we mentioned in Section 4. Thus, with essentially no restriction on p, we can ask: is there a way to construct the modular curve $X(p)$ geometrically from the invariant 3-tensor Θ? We will answer this question in the affirmative.

Suppose that instead of wanting to construct one hypersurface from another, we want to construct an invariant algebraic locus L in $\mathbb{P}^n$ from an invariant hypersurface $f = 0$, where f is an invariant. Here, the ambient geometry is supposed to be defined by a representation ρ of a finite group G. Since the invariants of ρ separate orbits of G, the algebraic locus L is the intersection of all of the invariant hypersurfaces containing it. Therefore, we can find a finite number of homogeneous invariants $I_1, \ldots, I_s$ which define the locus L set theoretically. If we can show that each of the invariants I_j is obtained from a covariant of f, then we can feel safe in asserting that the locus L can be constructed geometrically (in the set theoretic sense) from $f = 0$. This motivates the following definition.

Definition 6.2. Let $H : f = 0$ be a hypersurface in a projective space $P^d(\mathbb{C})$ of dimension d and let Z be a subvariety of $P^d(\mathbb{C})$. We say that Z *can be constructed geometrically from* H if the ideal defining Z is generated by covariants of f. We say that Z *can be constructed geometrically from* H *in the set theoretic sense* if Z is the set theoretic intersection of covariants of f.

From a theoretical point of view, the notion of geometric constructibility we are using is much too restrictive. If X is a G invariant hypersurface, it requires the locus to be an intersection of G-invariant hypersurfaces. While this may be true set theoretically, contemporary algebraic geometry requires us to consider the locus from a scheme theoretic point of view and it is certainly not reasonable to require the ideal defining Z to be generated by invariants. For example, if we take Z to be the singular locus of X, the ideal defining Z will be generated by the first partial derivatives of the form f defining X. Even if f is an invariant of G, the first partials of f in general will not be. Thus, passing to the singular locus of something geometrically constructible is not geometrical according to the definition we used. We could try to expand the notion by throwing in the singular locus construction, but that is arbitrary. It would be better to have a philosophical and comprehensive notion which is at the same time practical. Meanwhile, in the absence of one, I will leave things as they stand for the moment. It is rather like confining oneself to straightedge and compass constructions even though one cannot use them to trisect angles.

A second objection is that it our definition only addresses the question of constructing Z from a hypersurface X. It says nothing about constructing Z from some other locus W.

Remark 6.3. In Definition 6.2, there is no reason to confine ourselves to hypersurfaces except to preserve the geometric language. We could just as well speak of Z as being constructed from f. Since a polynomial is simply a symmetric tensor and the symmetric tensors form an irreducible representation of $\mathrm{SL}_{d+1}(\mathbb{C})$, we could instead take any irreducible representation of $\mathrm{SL}_{d+1}(\mathbb{C})$ on a finite-dimensional vector space W and choose an element Ξ of W. We can then consider covariants of Ξ and modify Definition 6.2 to speak of a subvariety Z of $\mathbb{P}^d$ being constructed geometrically from Ξ. We will use this more general definition in Theorem 7.7 below in the cases where -2 is not a square in $\mathbb{F}_q$.

7. Applications of Contemporary Invariant Theory

From the notion of geometric construction we are using, we see that it involves the notion of being able to extend mappings equivariant for one group to mappings equivariant for a larger group. Fortunately, contemporary invariant theory has been concerned with such questions.

We begin with a simple result whose statement and proof were kindly communicated to me by Gerry Schwarz.

THEOREM 7.1. *Let k be an algebraically closed field of characteristic zero. Let $\mathcal{G}$ be a reductive algebraic group and let V and W be representation spaces for $\mathcal{G}$ over k. Let x be a point of V such that the orbit $\mathcal{G} \cdot x$ of x under $\mathcal{G}$ is closed in V and let y be a point of W. Let $\mathcal{G}_x$ and $\mathcal{G}_y$ be the subgroups of $\mathcal{G}$ fixing x and y respectively. Then the following conditions are equivalent*:

(1) $\mathcal{G}_x \subseteq \mathcal{G}_y$;
(2) *there is a $\mathcal{G}$ equivariant polynomial mapping $\Gamma : V \to W$ such that $\Gamma(x) = y$.*

PROOF. The necessity of the condition is obvious. We prove the sufficiency. Since $\mathcal{G}_x \subseteq \mathcal{G}_y$, the map sending an element g of $\mathcal{G}$ to the element $g \cdot y$ of W factors through $\mathcal{G}/\mathcal{G}_x$. Since $X = \mathcal{G} \cdot x$ is closed in V, we may interpret the map as a $\mathcal{G}$-equivariant map f of X to W which sends x to y. The map extends to a morphism F of V to W. If N is sufficiently large, the space P of polynomial maps of degree $\leq N$ from V to W contains F and restriction to X maps P linearly and $\mathcal{G}$-equivariantly onto a space Q of maps from X to W containing f. Since $\mathcal{G}$ fixes f and since $\mathcal{G}$ is reductive, it follows that we can find an element of P which restricts to f and which is invariant under $\mathcal{G}$. □

REMARK 7.2. If $x = 0$ then the constant mapping with value y from V to W is homogeneous (of degree 0, if $y \neq 0$). If $y = 0$, then again the constant mapping with value y works. However, Theorem 7.1 does not let us conclude in general that we can find a *homogeneous* polynomial mapping Γ of V to W such that $\Gamma(x) = y$. The following counterexample is due to David Vogan. Let $\mathcal{G}$ be the a group of order 2, let V be the complex numbers $\mathbb{C}$ with $\mathcal{G}$ acting by ± 1 (that is, the non-trivial one-dimensional representation) and let W be $\mathbb{C}^2$ with the nontrivial element of $\mathcal{G}$ acting by interchange of coordinates (that is, the regular representation). Let $x = 1$ and let $y = (a, b)$, where $a \neq \pm b$. Then $\mathcal{G}_x = \mathcal{G}_y$ has order 1, so by Theorem 7.1, we can find a $\mathcal{G}$ equivariant mapping Γ from V to W carrying x to y. Suppose Γ is homogeneous. Then Γ must be of the form

$$\Gamma(z) = (az^m, bz^m)$$

for some nonnegative integer m. As z runs over all complex numbers, so does z^m, so the image of Γ will be the line in W generated by (a, b). Since V is invarant under $\mathcal{G}$, that line must be also. However, there are only two invariant lines in W and our hypothesis $a \neq \pm b$ implies that (a, b) doesn't lie on either of them.

That proves that Γ cannot be homogeneous. Close examination of this example leads to the additional condition that must be satisfied in order to guarantee the existence of a homogeneous mapping. This result, due to Dave Vogan, will be presented in Theorem 7.3 below. The proof given is also due to Vogan.

THEOREM 7.3 (VOGAN). *Let $\mathcal{G}, V, W, x, y$ be as in Theorem 7.1 and assume that x, y are both nonzero. Then the mapping Γ of Theorem 7.1 can be taken to be homogeneous if and only if the stabilizer $\mathcal{G}_{\mathbb{C}x}$ of the line $\mathbb{C}x$ through x is contained in the stabilizer $\mathcal{G}_{\mathbb{C}y}$ of the line $\mathbb{C}y$ through y.*

PROOF. Suppose that Γ is a $\mathcal{G}$ equivariant homogeneous mapping from V to W such that $\Gamma(x) = y$, say, homogeneous of degree m. Let z be a complex variable. As z runs over all complex numbers, so does z^m. Since $\Gamma(zx) = z^m y$, we conclude that the line through x is mapped by Γ onto the line through y. Since Γ is equivariant, if $g \in \mathcal{G}$ leaves the line through x invariant, it therefore must also leave the line through y invariant. This proves the necessity.

Next we assume the condition and prove its sufficiency. The group $\mathcal{G}_x$ is a normal subgroup of the group $\mathcal{G}_{\mathbb{C}x}$ and we denote the factor group by Z_x. We may identify the group Z_x with the multiplicative group of all nonzero complex numbers z such that zx lies in the orbit $\mathcal{G}\cdot x$ of x under $\mathcal{G}$. Similarly, we define the group Z_y. Both of the groups Z_x and Z_y are Zariski closed in the multiplicative group of $\mathbb{C}$. In particular, Z_x is either the whole multiplicative group or else it is a finite cyclic group. But it cannot be all of $\mathbb{C}^\times$ since we have assumed that the orbit of x is closed. Therefore the group Z_x is finite, say, of order m. Let ζ_m be a primitive m-th root of unity and let $g \in \mathcal{G}_{\mathbb{C}x}$ be such that $gx = \zeta_m x$. By hypothesis, $g \cdot y$ is a multiple of y, say λy. Since g^m fixes x, it must also fix y, so $\lambda^m = 1$. Therefore, $\lambda = \zeta_m^d$ for some integer d which is determined modulo m. It follows that any equivariant polynomial from V to W carrying x to y must be a sum of homogenous terms of degrees congruent to d modulo m. We can identify the affine coordinate ring of $\mathbb{C}x$ with the polynomial ring $\mathbb{C}[z]$ by means of the isomorphism $z \mapsto zx$ of $\mathbb{C}$ onto $\mathbb{C}x$. Denote by R_x the restriction to $\mathbb{C}x$ of the ring of invariants of $\mathcal{G}$, identified with a subring of $\mathbb{C}[z]$. Then R_x is generated by certain powers z^{rm} of z^m. By hypothesis, the orbit $\mathcal{G}\cdot x$ is closed. Since the group $\mathcal{G}$ is reductive, the invariants separate closed orbits. Therefore, the greatest common divisor of the integers rm is m. Since R_x is a ring, it must therefore contain z^{rm} for all sufficiently large values of r. In other words, for all sufficiently large r, there is an invariant j_r of degree rm such that $j_r(x) = 1$. Now let Γ be as in Theorem 7.1 and write Γ as a sum of its homogeneous parts:

$$\Gamma = \sum_{s=1}^{k} \Gamma_{d+r_s m},$$

where Γ_i denotes a homogeneous polynomial of degree i. Each of the Γ_i is of course $\mathcal{G}$ equivariant. Now choose r to be a sufficiently large integer and let

$$\Gamma' = \sum_{s=1}^{k} \jmath_{r-r_s}\Gamma_{d+r_s m}.$$

Then Γ' is equivariant, homogeneous of degree $d+rm$ and carries x to y. This proves the sufficiency. □

In order to apply Theorems 7.1 and 7.3, we need to have simple criteria for an orbit to be closed. The following theorem of Luna provides such a criterion.

THEOREM 7.4 [Luna 1975, p. 231]. *Let k be an algebraically closed field of characteristic* 0. *Let $\mathcal{G}$ be a reductive algebraic group over k and let $\mathcal{H}$ be a reductive subgroup of $\mathcal{G}$, not necessarily connected. Let $N_{\mathcal{G}}(H)$ denote the normalizer of $\mathcal{H}$ in $\mathcal{G}$. Then the following two conditions are equivalent*:

(1) *the group $N_{\mathcal{G}}(\mathcal{H})/\mathcal{H}$ is finite*;
(2) *in every rational representation of finite dimension $\mathcal{G} \to GL(M)$, the $\mathcal{G}$-orbit of any fixed point of $\mathcal{H}$ in M is closed in M.*

If r, s are nonnegative integers and M is a complex vector space, denote by $\bigotimes^{r,s} M$ the tensor product $M^{\otimes r} \otimes (M^*)^{\otimes s}$ viewed as a $GL(M)$ module. If $\mathcal{G}$ is a subgroup $GL(M)$, we call a $\mathcal{G}$ submodule of $\bigotimes^{r,s} M$ an (r,s) tensor module of $\mathcal{G}$. We will call a $\mathcal{G}$ module a *tensor module* if it is isomorphic to an (r,s) tensor module of $\mathcal{G}$ for some (r,s). An element of $\bigotimes^{r,s} M$ is called a *mixed tensor* of type (r,s) of M.

THEOREM 7.5. *Let G be a reductive algebraic group and let $\rho : G \to \mathrm{SL}(M)$ be a unimodular representation of G on a finite-dimensional complex vector space. Assume that $\rho(G)$ has finite index in its normalizer in* $\mathrm{SL}(M)$. (*This will be the case, e.g., if $\rho(G)$ is a maximal algebraic subgroup of* $\mathrm{SL}(M)$ *modulo scalars.*) *Let T_1, T_2 be mixed tensors on M, with T_1 of type (r,s) and T_2 of type (u,v). Assume that the isotropy group of T_2 in* $\mathrm{SL}(M)$ *contains G and that the isotropy group of T_1 in* $\mathrm{SL}(M)$ *coincides with $\rho(G)$ modulo scalars and is not all of* $\mathrm{SL}(M)$. *Then the following two conditions are equivalent*:

(1) *There exists a homogeneous* $\mathrm{SL}(M)$*-equivariant polynomial $\Gamma : \bigotimes^{r,s} M \to \bigotimes^{u,v} M$ such that $\Gamma(T_1) = T_2$.*
(2) *$u-v$ is a multiple of* $\gcd(r-s, m)$, *where m is the dimension of M.*

PROOF. If we take $\mathcal{G} = \mathrm{SL}(M)$ and $\mathcal{H} = \rho(G)$ in Luna's Theorem, the assumption on the normalizer of $\mathcal{H}$ implies that the $\mathrm{SL}(M)$ orbit of T_1 is closed in $\bigotimes^{r,s} M$. Next, we let $\mathcal{G} = \mathrm{SL}(M)$, $V = \bigotimes^{r,s} M$, $W = \bigotimes^{u,v} M$, $x = T_1$, $y = T_2$ in Theorems 7.1 and 7.3 and consider the hypotheses of these theorems. If $\mathcal{G}_x \subseteq \mathcal{G}_y$, our assumptions on the isotropy groups of T_1 and T_2 imply that $\mathcal{G}_{\mathbb{C}x} \subseteq \mathcal{G}_{\mathbb{C}y}$. This shows that if $\mathcal{G}, V, W, x, y$ satisfy the conditions of Theorem 7.1, they also satisfy the additional condition of Theorem 7.3 and that the $\mathcal{G}$ equivariant polynomial mapping Γ can therefore be taken to be homogeneous. Since $\rho(G) \subseteq \mathcal{G}_y$, the condition that $\mathcal{G}_x \subseteq \mathcal{G}_y$ is equivalent to the condition that every scalar in $\mathcal{G}_x$ lies in $\mathcal{G}_y$. Let z be a scalar multiplication in $\mathrm{SL}(M)$. Then z is an m-th root of unity, where m is the dimension of M. Under the action of z on V (or W), the tensor x (or y) is multiplied by z^{r-s} (or z^{u-v}, respectively). Therefore, $\mathcal{G}_x$ is generated by G and the d-th roots of unity, where $d = \gcd(m, r{-}s)$. Therefore, condition (1) is equivalent to the assertion that $u - v$ is a multiple of d, which is condition (2). This proves the theorem. □

Corollary 7.6. *Let $q = p^r$ be an odd prime power, where p is a prime number and where $r > 1$ if $p = 3$. Assume that $q \geq 11$ and $q \neq 13$. Let Θ be the unique invariant 3-tensor for $\mathrm{SL}_2(\mathbb{F}_q)$ on V_ν^ε. If q is not congruent to 1 modulo 6 then every invariant of $\mathrm{SL}_2(\mathbb{F}_q) \cdot \mathrm{Aut}(\mathbb{F}_q)$ on V_ν^ε arises from a covariant of Θ. If q is congruent to 1 modulo 6 then every invariant of degree divisible by 3 on V_ν^ε arises from Θ.*

Proof. This follows at once from Theorems 7.3 and 7.4 and and from the fact (see [Adler 1994]) that the group $\rho_\nu^\varepsilon(\mathrm{SL}_2(\mathbb{F}_q) \cdot \mathrm{Aut}(\mathbb{F}_q))$ is the precise automorphism group of Θ modulo scalars. □

Theorem 7.7. *Assume that $\eta = 1$ and let $n = (q - \varepsilon)/2$. Then there is a cubic contravariant of n-ary cubics which does not vanish on Θ. In particular, there **exists** a nonzero cubic contravariant of n-ary cubics.*

Proof. This follows at once from Theorem 7.5. Alternatively, by Theorem 7.4, the orbit of Θ is closed. In Theorem 7.1, let G be $\mathrm{SL}_n(\mathbb{C})$, V be the space of n-ary cubics, W be the dual space of V, x be Θ and let y be the differential operator D_Θ. By Theorem 7.1 there is a polynomial mapping λ of V into W which maps x onto y. Write λ as the sum $\lambda_0 + \lambda_1 + \cdots$ of its homogeneous components. Then each component is $\mathrm{SL}_n(\mathbb{C})$ invariant and is therefore a cubic contravariant of Θ. Since $\lambda(\Theta) = D_\Theta$, one of the terms $\lambda_i(\Theta)$ must be nonzero. Since any representation of $\mathrm{SL}_2(\mathbb{F}_q)$ of degree n has a unique cubic invariant (up to a scalar), we conclude that $\lambda_i(\Theta)$ must be a scalar multiple of D_Θ. Multiplying λ_i by the reciprocal of that scalar we obtain a cubic contravariant of n-ary cubics whose value on Θ is D_Θ. □

Theorem 7.8. *Suppose q is a an odd prime which is ≥ 11 and $\neq 13$. Then the modular curve $X(q)$ may be constructed geometrically from the 3-tensor Θ in the set theoretic sense. More precisely, if $\varepsilon = 1$ then the A-curve may be constructed from the restriction $\Theta|V^+$ of Θ to V^+, while if $\varepsilon = -1$, the z-curve may be constructed from the restriction $\Theta|V^-$ of Θ to V^-.*

Proof. Note that the invariants of $\mathrm{SL}_2(\mathbb{F}_q)$ separate orbits of $\mathrm{SL}_2(\mathbb{F}_q)$. It follows that the modular curve is the set theoretic intersection of all of the invariant hypersurfaces containing it. To prove the theorem, it therefore suffices to show that every $\mathrm{SL}_2(\mathbb{F}_q)$ invariant hypersurface Z arises set theoretically as a covariant of Θ. For set theoretic purposes, we may replace any invariant by its cube. Therefore, we can assume that the degree of the form F defining Z is divisible by 3. The theorem now follows from Corollary 7.6. □

8. Appendix: The Fundamental Intertwining Operator

In this section, we recall the Weil representation of a finite symplectic group and discuss a certain intertwining operator introduced in [Adler and Ramanan 1996]. We begin by recalling some of the notation of the fundamental paper [Weil 1964] and some of our own modifications of it.

8.1. Throughout this section, G will denote a locally compact abelian group, G^* its dual group and $\mathbb{T}$ the multiplicative group of all complex numbers of absolute value 1. The natural pairing between elements $g \in G$ and $g^* \in G^*$ is denoted $\langle g, g^* \rangle$ and the operation in G^* is also written additively. We will also assume that multiplication by 2 is an automorphism of G. Weil defines the group $A(G)$ to be the set $G \times G \times \mathbb{T}$ with the group law defined by

$$(g_1, g_1^*, t_1)(g_2, g_2^*, t_2) = (g_1 + g_2, g_1^* + g_2^*, \langle g_1, g_2^* \rangle t_1 t_2).$$

The group $A(G)$ is a locally compact topological group with the product topology. Its center is $\mathbb{T}$.

8.2. We note that the same construction will define a group if $\mathbb{T}$ is replaced by any subgroup $\mathbb{T}_0$ of $\mathbb{T}$ containing all of the values $\langle g, g^* \rangle$ with $g \in G$ and $g^* \in G^*$. Thus, we obtain a group which we denote $A_0(G)$ whose underlying set is $G \times G^* \times \mathbb{T}_0$. If $\mathbb{T}_0$ is not all of $\mathbb{T}$, we give $\mathbb{T}_0$ the discrete topology, so that $A_0(G)$ is likewise a locally compact topological group. Its center is $\mathbb{T}_0$. In practice, we will take $\mathbb{T}_0$ to be the smallest subgroup containing all of the values $\langle g, g^* \rangle$. In our applications, we will often find it better to work with the group $A_0(G)$ rather than $A(G)$.

8.3. Weil constructs certain automorphisms of $A(G)$ which induce the identity on the center $\mathbb{T}$ of $A(G)$. They are as follows.

8.3.1. *The automorphism $t_0(f)$ of $A(G)$.* By a *second degree character* of a locally compact abelian group H, we will mean a function f from H in to the circle group $\mathbb{T}$ such that the mapping $\beta : H \times H \to \mathbb{T}$, given by

$$\beta(g, h) = \frac{f(g+h)}{f(g)f(h)},$$

is a character of H in each variable separately. Thus, a second degree character is analogous to a quadratic polynomial without constant term. Indeed, if $p : \mathbb{R}^n \to \mathbb{R}$ is such a quadratic polynomial then the function $exp(2\pi i p)$ is a quadratic character.

If f is a second degree character of G, denote by $\rho : G \to G^*$ the associated symmetric morphism defined by

$$\langle g, h\rho \rangle = \frac{f(g+h)}{f(g)f(h)}.$$

Note that the homomorphism ρ is written on the right in Weil's notation. The automorphism $t_0(f)$ of $A(G)$ is defined by

$$t_0(f)(g, g^*, t) = (g, g\rho + g^*, f(g)t).$$

In practice, we will be concerned with the case in which f is even, that is, $f(g) = f(-g)$ for all $g \in G$. Since multiplication by 2 is assumed to be an automorphism of G, one can show that an even second degree character f is of the form

$$f(g) = \langle g/2, g\rho \rangle$$

where $g/2$ denotes the unique element of G such that $g/2 + g/2 = g$ and where ρ, as above, is the symmetric morphism associated to f.

8.3.2. *The automorphism $d_0(\alpha)$.* If α is a continuous automorphism of G, the automorphism $d_0(\alpha)$ of $A(G)$ is defined by

$$d_0(\alpha)(g, g^*, t) = (g\alpha, g^*\alpha^{*-1}, t).$$

Here α^* is the automorphism of G^* defined by composition with α, that is,

$$\langle g\alpha, g^* \rangle = \langle g, g^*\alpha^* \rangle$$

for all $g \in G$ and all $g^* \in G^*$.

8.3.3. *The automorphism $d_0'(\alpha)$.* Let $\gamma : G^* \to G$ be an isomorphism. We will be dealing with self-dual groups G, so this construction will not be empty. The automorphism $d_0'(\gamma)$ of $A(G)$ is defined by

$$d_0'(\gamma)(g, g^*, t) = (g^*\gamma, -g\gamma^{*-1}, \langle g, -g^* \rangle t).$$

Here $\gamma^* : G^* \to G$ is the isomorphism defined by

$$\langle g^*\gamma^*, h^* \rangle = \langle h^*\gamma, g^* \rangle$$

for all $g^*, h^* \in G^*$.The reader can easily verify that each of these automorphisms leaves the group $A_0(G)$ invariant and induces an automorphism on $A_0(G)$.

8.4. Weil denotes by $B(G)$ the group of all continuous automorphisms of $A(G)$. Each such automorphism induces an automorphism of the center $\mathbb{T}$ of $A(G)$. Such an automorphism must either be the identity on $\mathbb{T}$ or else must induce on $\mathbb{T}$ the automorphism $t \mapsto t^{-1}$. Weil denotes by $B_0(G)$ the subgroup of $B(G)$ inducing the identity automorphism on the subgroup $\mathbb{T}$. It is then easy to see that the elements of $B(G)$ and of $B_0(G)$ actually leave invariant the group $A_0(G)$ and induce automorphisms on it. Elements of $B_0(G)$ are uniquely determined by their restrictions to $B_0(G)$. Furthermore, any automorphism of $A_0(G)$ inducing the identity on $\mathbb{T}_0$ extends uniquely to an element of $B_0(G)$. So we will be free to identify $B_0(G)$ with the group of automorphisms of $A_0(G)$ inducing the identity on the center $\mathbb{T}_0$. It is also useful to consider certain endomorphisms of $A(G)$, and we will do so before we state Lemma 8.17.

8.5. The group $A(G)$ has a canonical unitary representation on the Hilbert space $L_2(G)$ of square integrable functions on G with respect to a Haar measure on G. That representation, denoted U, is defined as follows: if (g, g^*, t) is an element of $A(G)$ and if Φ is a square integrable function on G, then the function $\Phi' = U(g, g^*, t)\Phi$ is given by

$$\Phi'(x) = t\langle x, g^*\rangle \Phi(x + g)$$

for all $t \in G$. This notation U is slightly at variance with Weil's notation, according to which Φ' would be $tU(g, g^*)\Phi$. We find this modification of Weil's notation useful for our purposes. The representation U of $A(G)$ restricts to a representation of $A_0(G)$ on $L_2(G)$. We also denote that restriction by U.

8.6. Weil denotes by $\mathbf{A}(G)$ the image of $A(G)$ under the representation U and remarks that U induces a topological isomorphism of $A(G)$ onto $\mathbf{A}(G)$ where the latter is given the strong operator topology. Weil is consistent in the use of boldface fonts for operator versions of the groups and elements we have constructed. He denotes by $\mathbf{B}_0(G)$ the normalizer of $\mathbf{A}(G)$ in the group of all unitary operators on $L_2(G)$. He shows that there is a continuous homomorphism $\pi : \mathbf{B}_0(G) \to B_0(G)$, called the *canonical projection*, such that for all $S \in \mathbf{B}_0(G)$ and all $(g, g^*, t) \in A(G)$, we have

$$U(\pi(S)(g, g^*, t)) = S^{-1}U(g, g^*, t)S.$$

The kernel of π is the group of scalar multiplications by complex numbers of absolute value 1, which we may identify with the group $\mathbb{T}$.

8.7. Weil shows how to find elements of $\mathbf{B}_0(G)$ lying over elements of $B_0(G)$. This is in fact one of the important themes of [Weil 1964]. In the case of the elements $t_0(f)$, $d_0(\alpha)$ and $d_0'(\gamma)$ mentioned above, he gives the following elements of $\mathbf{B}_0(G)$ mapped to them respectively under the canonical projection π.

8.7.1. *The operator* $\boldsymbol{t}_0(f)$. Let f be a second degree character of G. Then the operator $\boldsymbol{t}_0(f)$ on $L_2(G)$ is defined by

$$(\boldsymbol{t}_0(f)\Phi)(x) = f(x)\Phi(x)$$

for all $\Phi \in L_2(G)$ and all $x \in G$.

8.7.2. *The operator* $\boldsymbol{d}_0(\alpha)$. Let α be an automorphism of G. Then the operator $\boldsymbol{d}_0(\alpha)$ is defined by

$$(\boldsymbol{d}_0(\alpha)\Phi)(x) = |\alpha|^{\frac{1}{2}}\Phi(x\alpha)$$

for all $\Phi \in L_2(G)$ and all $x \in G$, where $|\alpha|$ denotes the modulus of the automorphism α of the locally compact abelian group G. If X is a subset of G of finite positive measure, then the ratio of the measure of $X\alpha$ to the measure of X is $|\alpha|$. In case the group G is compact, we can take the set X to be all of G and we conclude that $|\alpha| = 1$ for a compact group.

8.7.3. *The operator $\boldsymbol{d}_0'(\gamma)$.* Let $\gamma : G^* \to G$ be an isomorphism. We define the operator $\boldsymbol{d}_0'(\gamma)$ on $L_2(G)$ by

$$(\boldsymbol{d}_0'(\gamma)\Phi)(x) = |\gamma|^{-\frac{1}{2}}\Phi^*(-x{\gamma^*}^{-1})$$

for all $\Phi \in L_2(G)$ and all $x \in G$. Here Φ^* is the Fourier transform of Φ, defined by

$$\Phi^*(x^*) = \int_G \Phi(x)\cdot\langle x, x^*\rangle \cdot dx$$

for all $x^* \in G^*$, the integral being taken with respect to a Haar measure dx on G. Thus the definition of Φ^* depends on the choice of dx, which is only unique up to a positive real factor. Specifically, if dx is replaced by cdx for some positive real number c, the value of Φ^* is likewise multiplied by c. Once one has chosen a Haar measure on G, there is canonically associated to it a Haar measure dx^* on G^*, called the dual measure, characterized by the relation

$$\int_G |\Phi(x)|^2 dx = \int_{G^*} |\Phi^*(x)|^2 dx^*$$

for all $\Phi \in L_2(G)$. Having chosen the Haar measure dx, we can therefore consider the modulus $|\gamma|$ of the isomorphism $\gamma : G^* \to G$. If X is a measurable subset of G^* with finite positive measure, $|\gamma|$ is the ratio of the dx-measure of $X\gamma$ to the dx^*-measure of X. If the Haar measure dx is replaced by cdx, where c is a positive real, then the value of $|\gamma|$ is multiplied by c^2. Thus, one sees that although both Φ^* and $|\gamma|$ depend on the choice of dx, the definition of the operator $\boldsymbol{d}_0'(\gamma)$ does not.

8.8. The reader can verify that the canonical projection maps $\boldsymbol{t}_0(f)$, $\boldsymbol{d}_0(\alpha)$ and $\boldsymbol{d}_0'(\gamma)$ respectively to $t_0(f)$, $d_0(\alpha)$ and $d_0'(\gamma)$.

8.9. Weil also showed that in general there is no homomorphism from $B_0(G)$ to $\mathbf{B}_0(G)$ whose composition with the canonical projection is the identity on $B_0(G)$. However for certain subgroups[1] of $B_0(G)$ which he denotes $B_0(G,\Gamma)$, where Γ is a closed subgroup of G, he was able to define canonical homomorphisms from $B_0(G,\Gamma)$ to $\mathbf{B}_0(G)$ whose composition with the canonical projection is the identity on $B_0(G,\Gamma)$. In the special case where G is the adèle group of a vector space of finite dimension over an **A**-field (that is, a number field or an algebraic function field in one variable over a finite field), the representation so obtained is commonly known as the *Weil representation.* In his general setting, Weil denoted his homomorphism from $B_0(G,\Gamma)$ by r_Γ. If an element of $B_0(G)$ is expressed as a product of elements of the form $t_0(f)$, $d_0(\alpha)$ and $d_0'(\gamma)$, one can take the product of the operators $\boldsymbol{t}_0(f)$, $\boldsymbol{d}_0(\alpha)$ and $\boldsymbol{d}_0'(\gamma)$ associated to these elements to obtain an element of $\mathbf{B}_0(G)$. This element depends, however, on the manner

[1]The subgroup $B_0(G,\Gamma)$ of $B_0(G)$ consists of all elements of $B_0(G)$ which leave invariant the subgroup $\Gamma\times\Gamma_*\times\mathbb{T}$ of $A(G)$, where Γ_* denotes the annihilator of Γ in G^*.

in which one writes the given element of $B_0(G)$ as a product of elements of the form $t_0(f)$, $d_0(\alpha)$ and $d_0'(\gamma)$.

8.10. Following Weil, we denote by $\mathrm{Sp}(G)$ the symplectic group of G, which by definition is the group of all continuous automorphisms of $G\times G^*$ preserving the alternating bicharacter of $G\times G^*$ given by

$$((g_1,g_1^*),(g_2,g_2^*))\mapsto\frac{\langle g_1,g_2^*\rangle}{\langle g_2,g_1^*\rangle}.$$

Following [Adler 1989], we denote by $\mathrm{Sp}'(G)$ the centralizer of $d_0(-1_G)$ in $B_0(G)$. Since every element of $B_0(G)$ leaves $\mathbb{T}$ invariant, every such element induces an automorphism of $A(G)/\mathbb{T}$. The latter group is isomorphic to $G\times G^*$ and the induced automorphism is in fact symplectic. Therefore we have a natural homomorphism from $B_0(G)$ to $\mathrm{Sp}(G)$ and it is not difficult to show using [Weil 1964, § 5] that $\mathrm{Sp}'(G)$ is mapped isomorphically onto $\mathrm{Sp}(G)$.

8.11. In [Adler 1989], the group $\mathbf{B}_1(G)$ is defined to be the subgroup of $\mathbf{B}_0(G)$ consisting of all operators in $\mathbf{B}_0(G)$ that commute with $\boldsymbol{d}_0(-1_G)$. In Lemma 25.2 of that paper, it is shown that the canonical projection maps $\mathbf{B}_1(G)$ surjectively onto $\mathrm{Sp}'(G)$ provided the following hypothesis holds:

(E) The group $\mathrm{Sp}(G)$ is generated by the elements $\left(\begin{smallmatrix}\alpha&\beta\\\gamma&\delta\end{smallmatrix}\right)$ of $\mathrm{Sp}(G)$ with $\gamma: G^*\to G$ an isomorphism.

Denote by V^+ and V^- respectively the 1 and -1 eigenspaces of $\boldsymbol{d}_0(-1_G)$ in $L_2(G)$. It is the same to say that V^+, V^- are respectively the spaces of *even* and *odd* functions in $L_2(G)$. Since the elements of $\mathbf{B}_1(G)$ commute with $\boldsymbol{d}_0(-1_G)$, they leave V^+ and V^- invariant. If S is an element of $\mathbf{B}_1(G)$, we will denote by S^+ and S^- respectively the operators induced by S on V^+ and V^-.

8.12. For the rest of this section, we will assume that G is a finite abelian group of odd order $2N+1$. In this case, $L_2(G)$ is finite-dimensional, V^+ has dimension $N+1$ and V^- has dimension N. According to [Adler 1989, Lemma 26.1], the mapping

$$S\mapsto\chi(S)=\frac{\det(S^+)}{\det(S^-)}$$

is a character of $\mathbf{B}_1(G)$ such that $\chi(t)=t$ for all $t\in\mathbb{T}$. Denote by ϕ the homomorphism from $\mathbf{B}_1(G)$ to itself given by

$$\phi(S)=\chi(S)^{-1}S.$$

The image of ϕ is denoted by $\mathrm{Sp}''(G)$. If G satisfies hypothesis (E) then [Adler 1989, Lemma 26.2] says that the canonical projection π maps $\mathrm{Sp}''(G)$ isomorphically onto $\mathrm{Sp}(G)$. The inverse of this isomorphism is denoted r'. The representation r' is also called the Weil representation.

8.13. According to [Adler 1989, Theorem 27.1], for any second degree character f of G such that $f(-x) = f(x)$ for all $x \in G$, we have

$$r'(t_0(f)) = \boldsymbol{t}_0(f).$$

So the Weil representation r' is given by Weil's explicit lifting in this case. This is not so, however, for the elements of $\mathrm{Sp}(G)$ of the form $d_0(\alpha)$ and $d_0'(\gamma)$.

8.14. For the rest of this section, we will further specialize G by assuming that G is actually a vector space of finite dimension n over the field $\mathbb{F}_p$ with p elements. Then hypothesis (E) holds in this case. Since G is compact, the factor $|\alpha|^{\frac{1}{2}}$ in $\boldsymbol{d}_0(\alpha)$ is always equal to 1. And with respect to the self-dual Haar measure on G, the factor $|\gamma|^{-\frac{1}{2}}$ in $\boldsymbol{d}_0'(\gamma)$ is always equal to 1. So we will disregard such factors in what follows. We will also use only the unique self-dual Haar measure on G. According to [Adler 1989, Theorem 27.4], for any automorphism α of G, we have

$$r'(d_0(\alpha)) = \left(\frac{\det(\alpha)}{p}\right) \boldsymbol{d}_0(\alpha),$$

where the first factor on the right side is the quadratic residue symbol of $\det(\alpha)$ in $\mathbb{F}_p$. As for the value of r' on elements of the form $d_0'(\gamma)$, let $\gamma : G^* \to G$ be an isomorphism, let ρ be a symmetric isomorphism of G onto G^* and let α be the automorphism of G defined by $\gamma = \rho^{-1}\alpha$. Then [Adler 1989, Theorem 27.5], we have

$$\chi(\boldsymbol{d}_0(\gamma)) = \gamma(f) \left(\frac{\det(\alpha)}{p}\right),$$

where f is the second degree character of G given by

$$f(x) = \langle x/2, x\rho\rangle$$

and

$$\gamma(f) = \int_G f dx$$

is the integral of f with respect to the unique self-dual Haar measure on G. Since hypothesis (E) holds, the elements of $\mathrm{Sp}(G)$ of the form $t_0(f)$, $d_0(\alpha)$ and $d_0'(\gamma)$ generate $\mathrm{Sp}(G)$. Therefore, the results just quoted amount to a complete determination of the Weil representation r'.

8.15. Let ν be an integer. Denote by σ_ν the mapping from $A(G)$ to itself given by

$$\sigma_\nu(g, g^*, t) = (\nu g, g^*, t^\nu).$$

Then one can verify directly that σ_ν is a continuous endomorphism of $A(G)$. Furthermore, it clearly leaves invariant the group $A_0(G)$. We are concerned with cases in which σ_ν actually induces an automorphism of $A_0(G)$; this happens, for example, if multiplication by ν is an automorphism of G and if $\mathbb{T}_0$ doesn't contain any ν-th roots of unity other than 1. When σ_ν induces an automorphism of $A_0(G)$, we will also denote that automorphism by σ_ν, or more simply by σ in case we don't need to indicate ν. It is then clear that conjugation by σ_ν induces an automorphism s_ν of $B_0(G)$. We will also write s instead of s_ν when we do not need to refer directly to the integer ν. The automorphism σ_ν of $A_0(G)$ depends only on the congruence class of ν modulo p and we will freely refer to ν as a congruence class modulo p instead of as an integer.

8.16. Let ν be a nonzero element of $\mathbb{F}_p$. Then multiplication by ν induces an automorphism of G. Accordingly, the automorphism s_ν of $B_0(G)$ is well defined. Furthermore, since s_ν itself commutes with $d_0(-1_G)$, s_ν actually induces an automorphism of $\mathrm{Sp}(G)$. That automorphism will also be denoted s_ν. The composition of s_ν with the Weil representation r' will be denoted r'_ν. Of course, the representation r' is the same as r'_1.

LEMMA 8.17. *Let f be an even second degree character of G. Let α be an automorphism of G and let $\gamma : G^* \to G$ be an isomorphism. Then*

$$r'_\nu(t_0(f)) = r'(t_0(f^\nu)),$$
$$r'_\nu(d_0(\alpha)) = r'(d_0(\alpha)),$$
$$r'_\nu(d'_0(\gamma)) = r'(d'_0(\gamma/\nu)).$$

PROOF. Let $\rho : G \to G^*$ be the symmetric isomorphism associated to f. Then

$$f(x) = \langle x/2, x\rho\rangle$$

for all $x \in G$. Let (g, g^*, t) be an arbitrary element of $A_0(G)$. Then

$$\sigma_\nu(g, g^*, t) = (\nu g, g^*, t^\nu).$$

Therefore

$$\begin{aligned}\sigma_\nu^{-1} t_0(f)\sigma_\nu(g, g^*, t) &= \sigma_\nu t_0(f)(\nu g, g^*, t^\nu)\\ &= \sigma_\nu^{-1}(\nu g, \nu g\rho + g^*, f(\nu g)t^\nu)\\ &= (g, g\nu\rho + g^*, f(g)^\nu t) = t_0(f^\nu)(g, g^*, t),\end{aligned}$$

since f^ν is the even second degree character associated to $\nu\rho$. Therefore,

$$r'_\nu(t_0(f)) = r'(s_\nu(t_0(f))) = r'(t_0(f^\nu)).$$

Since $d_0(\alpha)$ commutes with σ_ν, we similarly have

$$r'_\nu(d_0(\alpha)) = r'(s_\nu(d_0(\alpha))) = r'(d_0(\alpha)).$$

Finally,

$$\begin{aligned}\sigma_\nu^{-1} d'_0(\gamma)\sigma_\nu(g, g^*, t) &= \sigma_\nu^{-1} d'_0(\gamma)(\nu g, g^*, t^\nu)\\ &= \sigma_\nu^{-1}(g^*\gamma, -\nu g\gamma^{*\,-1}, \langle g, -g^*\rangle^\nu t^\nu)\\ &= (g^*\gamma/\nu, -\nu g\gamma^{*\,-1}, \langle g, -g^*\rangle t) = d'_0(\gamma/\nu),\end{aligned}$$

so $r'_\nu(d_0(\gamma)) = r'(d_0(\gamma/\nu))$. □

8.18. The representations r' and r'_ν of $\mathrm{Sp}(G)$ on $L_2(G)$ induce representations, also denoted respectively by r' and r'_ν, on the tensor powers of $L_2(G)$. Since we can identify $L_2(G)\otimes L_2(G)$ with $L_2(G\times G)$ canonically, in particular we obtain representations r', r'_ν of $\mathrm{Sp}(G)$ on $L_2(G\times G)$. Similarly, we will freely regard operators such as $\boldsymbol{t}_0(f)$, $\boldsymbol{d}_0(\alpha)$ and $\boldsymbol{d}'_0(\gamma)$ as operators on $L_2(G\times G)$. The identities of the preceding lemma therefore hold without modification when both sides are regarded as operators on $L_2(G\times G)$. It is useful to observe that when the number of tensor factors is a positive integer N, the factors such as $\left(\frac{\det(\alpha)}{p}\right)$ and $\gamma(f)$ are replaced by their N-th powers. Since each of these factors equals ±1, it follows that when the number of tensor factors is even, the factors become equal to 1. We will in fact be concerned with the case of two tensor factors, so we will not have to worry about these factors further. However, we also have occasion to consider tensor products of representations r'_ν for different ν. We will then use multi-indices for the subscript of r'. Explicitly, we will denote by $r'_{\mu*\nu}$ the tensor product of r'_μ and r'_ν, where μ and ν may be either single integers or multi-indices and where $*$ denotes concatenation of lists of integers.

8.19. We now introduce an operator that has proved to be of fundamental importance in our work. It is the operator $\mathfrak{T}$ on $L_2(G\times G)$ given by

$$(\mathfrak{T}\Phi)(x,y)=\Phi\left(\frac{x+y}{2},\frac{x-y}{2}\right)$$

for all $\Phi\in L_2(G\times G)$ and all $x,y\in G$. We will refer to the operator $\mathfrak{T}$ as the *fundamental intertwining operator*. We then have the following result, stated and proved in special cases in [Adler and Ramanan 1996] but undoubtedly well known, which justifies this terminology.

THEOREM 8.20. *The operator $\mathfrak{T}$ is an isomorphism between the representations r' and r'_2 on $L_2(G\times G)$.*

PROOF. We will simply verify this for the elements of $\mathrm{Sp}(G)$ of the form $t_0(f)$, $d_0(\alpha)$ and $d'_0(\gamma)$. Let $\Phi\in L_2(G\times G)$ and $x,y\in G$. Then

$$\begin{aligned}
(\mathfrak{T}r'_2(t_0(f))(\Phi))(x,y)&=(r'_2(t_0(f))(\Phi))\left(\frac{x+y}{2},\frac{x-y}{2}\right)\\
&=(r'(t_0(f^2))(\Phi))\left(\frac{x+y}{2},\frac{x-y}{2}\right)\\
&=f^2\left(\frac{x+y}{2}\right)f^2\left(\frac{x-y}{2}\right)\Phi\left(\frac{x+y}{2},\frac{x-y}{2}\right)\\
&=\left\langle\frac{x+y}{4},\frac{x+y}{2}2\rho\right\rangle\left\langle\frac{x-y}{4},\frac{x-y}{2}2\rho\right\rangle\Phi\left(\frac{x+y}{2},\frac{x-y}{2}\right)\\
&=\langle x/2,x\rho\rangle\langle y/2,y\rho\rangle\Phi\left(\frac{x+y}{2},\frac{x-y}{2}\right)\\
&=(r'(t_0(f))(\mathfrak{T}\Phi))(x,y),
\end{aligned}$$

which proves the theorem in the case of $t_0(f)$. For $d_0(\alpha)$ we simply have

$$\begin{aligned}(\mathcal{T}r_2'(d_0(\alpha))(\Phi))(x,y) &= (r_2'(d_0(\alpha))(\Phi))\Big(\frac{x+y}{2},\frac{x-y}{2}\Big)\\ &= (r'(d_0(\alpha))(\Phi))\Big(\frac{x+y}{2},\frac{x-y}{2}\Big)\\ &= \Phi\Big(\frac{x+y}{2}\alpha,\frac{x-y}{2}\alpha\Big)\\ &= (\mathcal{T}(\Phi))(x\alpha,y\alpha) = (r'(d_0(\alpha))\mathcal{T}(\Phi))(x,y),\end{aligned}$$

which proves the theorem in the case of $d_0(\alpha)$. Finally, if $\gamma : G^* \to G$ is an isomorphism and $\rho : G \to G^*$ is a symmetric isomorphism, we let $\alpha = \rho\gamma$. We then have

$$\begin{aligned}&(\mathcal{T}r_2'(d_0'(\gamma))(\Phi))(x,y)\\ &\quad= (r_2'(d_0'(\gamma))(\Phi))\Big(\frac{x+y}{2},\frac{x-y}{2}\Big)\\ &\quad= (r'(d_0'(\gamma/2))(\Phi))\Big(\frac{x+y}{2},\frac{x-y}{2}\Big)\\ &\quad= \int_{G\times G}\Phi(u,v)\Big\langle u,-\frac{x+y}{2}(\gamma/2)^{*-1}\Big\rangle\Big\langle v,-\frac{x-y}{2}(\gamma/2)^{*-1}\Big\rangle\,dx\,dy\\ &\quad= \int_{G\times G}\Phi\Big(\frac{a+b}{2},\frac{a-b}{2}\Big)\Big\langle\frac{a+b}{2},-(x+y)\gamma^{*-1}\Big\rangle\Big\langle\frac{a-b}{2},-(x-y)\gamma^{*-1}\Big\rangle\,dx\,dy\\ &\quad= \int_{G\times G}\Phi\Big(\frac{a+b}{2},\frac{a-b}{2}\Big)\langle a,-x\gamma^{*-1}\rangle\langle b,-y\gamma^{*-1}\rangle\,da\,db\\ &\quad= \int_{G\times G}(\mathcal{T}\Phi)(a,b)\langle a,-x\gamma^{*-1}\rangle\langle b,-y\gamma^{*-1}\rangle\,da\,db\\ &\quad= (r'(d_0'(\gamma))\mathcal{T}(\Phi))(x,y). \qquad \square\end{aligned}$$

8.21. The subspaces V^+ and V^- are invariant under r'_ν for all integers ν. The representations one obtains on these spaces for each ν will be denoted ρ_ν^+ and ρ_ν^- respectively. Although the spaces V^+ and V^- themselves do not depend on ν, we will denote them V_ν^+ and V_ν^- respectively whenever we wish to emphasize that one is acting on them via ρ_ν^+ and ρ_ν^-. We also note that, for the purposes of studying the relevant bicycles, all of the representations r'_ν of $\mathrm{Sp}_{2r}(\mathbb{F}_p)$ (resp. $\mathrm{SL}_2(\mathbb{F}_q)$) are equivalent under the group of automorphisms of $\mathrm{Sp}_{2r}(\mathbb{F}_p)$ (resp. $\mathrm{SL}_2(\mathbb{F}_q)$) and the representation $r'_{-\nu}$ is the dual of the representation r'_ν. The same remarks apply, *mutatis mutandis*, to the representations $\rho_\nu^\pm$. In particular, the rings of invariants of all of these representations form bicycles.

8.22. Parts of the following result, and the relevant principles for proving it, may be found in [Adler and Ramanan 1996, pp. 54, 55, 74].

LEMMA 8.23. *We have*

$$\mathcal{T}\big(\mathrm{Sym}^2(V_\nu^+)\big)=\mathrm{Sym}^2(V_{2\nu}^+),\qquad \mathcal{T}\big(\textstyle\bigwedge^2(V_\nu^-)\big)=\bigwedge^2(V_{2\nu}^-),$$
$$\mathcal{T}\big(\mathrm{Sym}^2(V_\nu^-)\big)=\textstyle\bigwedge^2(V_{2\nu}^+),\qquad \mathcal{T}\big(\bigwedge^2(V_\nu^+)\big)=\mathrm{Sym}^2(V_{2\nu}^-).$$

PROOF. Denote by α, β two numbers each of which is equal to ±1. Denote by $W(\alpha,\beta)$ the space of all complex valued functions $f(x,y)$ on $G\times G$ such that

$$f(y,x)=\alpha f(x,y)$$

and

$$f(-x,y)=\beta f(x,y).$$

We note that these two conditions imply

$$f(x,-y)=\beta f(x,y).$$

We then have

$$W(\ \ 1,1)=\mathrm{Sym}^2(V_{2\nu}^+),\qquad W(\ \ 1,-1)=\mathrm{Sym}^2(V_{2\nu}^-),$$
$$W(-1,1)=\textstyle\bigwedge^2(V_{2\nu}^+),\qquad W(-1,-1)=\bigwedge^2(V_{2\nu}^-).$$

If $f\in W(\alpha,\beta)$, then

$$(\mathcal{T}f)(y,x)=f\Big(\frac{y+x}{2},\frac{y-x}{2}\Big)=\beta f\Big(\frac{y+x}{2},\frac{x-y}{2}\Big)=\beta(\mathcal{T}f)(x,y)$$

and

$$(\mathcal{T}f)(-x,y)=f\Big(\frac{-x+y}{2},\frac{-x-y}{2}\Big)=f\Big(\frac{x-y}{2},\frac{x+y}{2}\Big)=\alpha f\Big(\frac{x+y}{2},\frac{x-y}{2}\Big).$$

This shows that $\mathcal{T}$ maps $W(\alpha,\beta)$ into $W(\beta,\alpha)$. Since $L_2(G\times G)$ is finite-dimensional and is the direct sum of the spaces $W(\alpha,\beta)$, we are done. □

8.24. We can generalize the fundamental intertwining operator in the following way. Let a,b be any elements of $\mathbb{F}_p$ such that $a^2+b^2\neq0$. Then we define the operator $\mathcal{T}_{a,b}$ on $L^2(G\otimes G)$ by the rule

$$(\mathcal{T}_{a,b}\Phi)(x,y)=\Phi(ax+by,-bx+ay).$$

A computation similar to the one in Theorem 8.20 shows that $\mathcal{T}_{a,b}$ is an intertwining operator between the representations r'_ν and $r'_{\nu(a^2+b^2)}$ on $L^2(G\times G)$, that is,

$$\mathcal{T}_{a,b}\circ r'_\nu=r'_{\nu(a^2+b^2)}\circ\mathcal{T}_{a,b}.$$

The fundamental intertwining operator is then $\mathcal{T}_{\frac12,\frac12}$. We do need the more general interwining operator $\mathcal{T}_{a,b}$ in Section 4, for example. It should be noted that the preceding lemma does not hold in general with $\mathcal{T}$ replaced by $\mathcal{T}_{a,b}$, but it does hold if $a=b$.

8.25. We can identify the group $\mathrm{Sp}(G)$ with the finite symplectic group $\mathrm{Sp}_{2s}(\mathbb{F}_p)$, where the order of G is p^{2s}. If $s = nr$, where n, r are positive integers, the group $\mathrm{Sp}_{2n}(\mathbb{F}_q)$, where $q = p^r$, is naturally a subgroup of $\mathrm{Sp}_{2s}(\mathbb{F}_p)$. The Weil representations of $\mathrm{Sp}(G)$ therefore give rise to representations of these two groups which we also refer to as Weil representations. We will also retain the notations r' and r'_ν.

8.26. By drawing on the methods and concepts of [Weil 1964, § 49], the results of this Appendix can be extended without difficulty to treat the Weil representation of the symplectic group $\mathrm{Sp}_{2n}(\mathbb{F}_q)$ directly. Instead of the group $A_0(G)$, one introduces the group $A^{\#}(G)$ whose underlying point set is $G \times G^* \times \mathbb{F}_q$ and whose operation is given by

$$(g_1, g_1^*, u_1)(g_2, g_2^*, u_2) = (g_1 + g_2,\ g_1^* + g_2^*,\ [g_1, g_2^*] + u_1 + u_2).$$

In order to explain the pairing $[\,\cdot\,,\cdot\,]$ that appears on the right, denote by tr the trace from $\mathbb{F}_q$ to $\mathbb{F}_p$ and by τ the character of the additive group of $\mathbb{F}_q$ given by

$$\tau(x) = \zeta_p^{\mathrm{tr}(x)}.$$

There is a natural structure on G^* of vector space over $\mathbb{F}_q$ induced by that of G. Denote by G' the dual space of the $\mathbb{F}_q$ vector space G. There is a canonical isomorphism of G' onto G^* which associates to an element λ of G' the composition $\tau \circ \lambda$ of λ and τ. By means of this isomorphism, we canonically identify G^* with G'. The pairing $[\,\cdot\,,\cdot\,]$ is then the natural pairing from $G \times G'$ to $\mathbb{F}_q$.

8.27. We identify $\mathrm{Sp}_{2n}(\mathbb{F}_q)$ with the centralizer in $\mathrm{Sp}(G)$ of all elements of the form $d_0(\alpha)$ where α is scalar multiplication by an element of $\mathbb{F}_q$. One can then verify that $\mathrm{Sp}_{2n}(\mathbb{F}_q)$ acts as a group of automorphisms of $A^{\#}(G)$ in a natural way. Explicitly, suppose an element β of $\mathrm{Sp}_{2n}(\mathbb{F}_q)$ acts on $A_0(G)$ by the rule

$$(w, t) \mapsto (w\beta, f(w)t),$$

where $w \in G \times G^*$ and where f is a second degree character of G such that $f(-x) = f(x)$ for all $x \in G$. Then as in [Weil 1964, § 49] we can write f uniquely in the form

$$f(x) = \tau(F(x)),$$

where $F : G \times G^* \to \mathbb{F}_q$ is a quadratic form on the $\mathbb{F}_q$ vector space $G \times G^*$. Then β acts on $A^{\#}(G)$ by

$$\beta(w, u) = (w\beta, u + F(w)).$$

8.28. If ν is a nonzero element of $\mathbb{F}_q$, we can define the automorphism σ_ν of $A^{\#}(G)$ by the rule

$$\sigma_\nu(g,g^*,u)=(\nu g,g^*,\nu u).$$

Conjugation by σ_ν leaves $\mathrm{Sp}_{2n}(\mathbb{F}_q)$ invariant and induces an automorphism on it which we denote s_ν. We denote by r'_ν the composition of the Weil representation r' with s_ν. Thus, we can obtain more Weil representations of $\mathrm{Sp}_{2n}(\mathbb{F}_q)$ in this way than we could using only integers ν prime to p. The importance of considering these more general Weil representations is that in case ν happens to be a square in $\mathbb{F}_q$, the Weil representation is equivalent to the original Weil representation. If $q=p^r$ with r even, then every element of $\mathbb{F}_p$ will be a square in $\mathbb{F}_q$ and we will get nothing new. But by taking ν to be an element of $\mathbb{F}_q$ which is not a square in $\mathbb{F}_q$, we do get something new.

8.29. In connection with the bicycles we are considering, note that even if ν is not a square in $\mathbb{F}_q$, the Weil representation r'_ν can be obtained from r' by composition with an automorphism of $\mathrm{Sp}_{2n}(\mathbb{F}_q)$.

8.30. With these preliminaries, the fundamental intertwining operator still behaves as described in Lemma 8.23 and the intertwining operators $\mathcal{T}_{a,b}$ can be defined more generally whenever a,b are elements of $\mathbb{F}_q$ such that $a^2+b^2\neq 0$.

8.31. In closing, we show that if -1 is a square in $\mathbb{F}_q$ then the representation of $\mathrm{Sp}_{2n}(\mathbb{F}_q)$ on V_ν^+ is orthogonal and on V_ν^- is symplectic. Since -1 is a square, let a be a square root of -1 in $\mathbb{F}_q$. Then the mapping τ_a from $L_2(\mathbb{F}_q)$ to itself given by

$$\tau_a\Phi(x)=\Phi(ax)$$

normalizes $r'(\mathrm{Sp}_{2n}(\mathbb{F}_q))$. Indeed, an easy direct computation shows that τ_a intertwines r'_ν and $r'_{-\nu}$. On the other hand, we have the canonical $\mathrm{Sp}_{2n}(\mathbb{F}_q)$ invariant pairing between r'_ν and $r'_{-\nu}$ given by

$$\langle\Phi_1,\Phi_2\rangle=\sum\Phi_1(x)\Phi_2(x),$$

where the summation runs over all $x\in\mathbb{F}_q^n$. We therefore obtain the bilinear pairing $\mathcal{Q}$ on $L_2(\mathbb{F}_q^n)$ given by

$$\mathcal{Q}(\Phi_1,\Phi_2)=\sum\Phi_1(x)\Phi_2(ax).$$

If Φ_1,Φ_2 are even functions on $\mathbb{F}_q^n$, we have

$$\mathcal{Q}(\Phi_2,\Phi_1)=\mathcal{Q}(\Phi_1,\Phi_2),$$

hence the restriction $\mathcal{Q}^+$ of $\mathcal{Q}$ to V_ν^+ is symmetric. One sees that it is also nonzero by taking $\Phi_1=\Phi_2$ to be the function which is 1 at 0 and 0 everywhere else. This shows that V_ν^+ is orthogonal and gives the invariant quadratic form explicitly as

$$\Phi\mapsto\sum\Phi(x)\Phi(ax).$$

We will also denote this quadratic form by $\mathcal{Q}^+$. On the other hand, if Φ_1, Φ_2 are both odd functions on $\mathbb{F}_q^n$, we have

$$\mathcal{Q}(\Phi_2, \Phi_1) = -\mathcal{Q}(\Phi_1, \Phi_2),$$

which shows that the restriction $\mathcal{Q}^-$ of $\mathcal{Q}$ to V_ν^- is alternating. To see that it is nonzero, let y be a nonzero element of $\mathbb{F}_q^n$ and let Φ_1 be 1 at y, -1 at $-y$ and 0 everywhere else and let $\Phi_2(x) = \Phi_1(ax)$ for all x. Then we have

$$\mathcal{Q}^-(\Phi_1, \Phi_2) = \sum \Phi_1(x)\Phi_2(ax) = -\sum \Phi_1(x)^2 = -2,$$

which proves that $\mathcal{Q}^-$ is also nonzero. Thus V_ν^- is symplectic and its invariant alternating form $\mathcal{Q}^-$ is given explicitly.

Acknowledgements

I am grateful to Gerry Schwarz and to David Vogan for their crucial help with the results of Section 7 and to Ofer Gabber for pointing out a simple oversight in the formulation of the Bicycle Conjecture. I am also thankful to IHES for lodging during the second half of 1994 and for support during the fall of 1994, as well as for allowing me to conduct an informal seminar that provided the crucial stimulation I needed to make final revisions in the manuscript. I am especially grateful to Mischa Gromov, Marcel Berger and Jean-Pierre Bourguignon for their sympathetic help in making the facilities of IHES available to me.

References

[Adler 1981] A. Adler, "On the ring of invariants of $\mathrm{PSL}_2(\mathbf{F}_{11})$ acting on $\mathbf{C}^5$", preprint, Tata Institute, Bombay, 1981.

[Adler 1989] A. Adler, "On the Weil representation", preprint, 1989. Reprinted as [Adler and Ramanan 1996, Appendix 1].

[Adler 1992a] A. Adler, "Cubic invariants for $\mathrm{SL}_2(\mathbf{F}_q)$", *J. Algebra* **145**:1 (1992), 178–186.

[Adler 1992b] A. Adler, "Invariants of $\mathrm{PSL}_2(\mathbf{F}_{11})$ acting on $\mathbf{C}^5$", *Comm. Algebra* **20**:10 (1992), 2837–2862.

[Adler 1994] A. Adler, "On the automorphism groups of certain hypersurfaces, II", *Comm. Algebra* **22**:7 (1994), 2319–2366.

[Adler 1997] A. Adler, "The Mathieu group M_{11} and the modular curve $X(11)$", *Proc. London Math. Soc.* (3) **74**:1 (1997), 1–28.

[Adler and Ramanan 1996] A. Adler and S. Ramanan, *Moduli of abelian varieties*, Lecture Notes in Math. **1644**, Springer, Berlin, 1996.

[Baker 1935] H. F. Baker, "Note introductory to the study of Klein's group of order 168", *Proc. Camb. Phil. Soc.* **31** (1935), 168–181.

[Burckhardt 1893] H. Burckhardt, "Untersuchungen aus dem Gebiete der hyperelliptischen Modulfunktionen III", *Mathematische Annalen* **41** (1893), 313.

[Coble 1917] A. Coble, "Point sets and allied Cremona groups, III", *Trans. Amer. Math. Soc.* **18** (1917), 331–372.

[Dieudonné and Carrell 1971] J. A. Dieudonné and J. B. Carrell, *Invariant theory, old and new*, Academic Press, New York, 1971. Alternatively see article of the same name in *Advances in Math.* **4** (1970), 1–80.

[Edge 1947] W. L. Edge, "The Klein group in three dimensions", *Acta Math.* **79** (1947), 153–223.

[Grace and Young 1903] J. H. Grace and A. Young, *The algebra of invariants*, University Press, Cambridge, 1903. Reprinted by G. E. Stechert, New York, 1941 and by Chelsea, Bronx, NY, 1965.

[Klein 1879a] F. Klein, "Ueber die Transformationen siebenter Ordnung der elliptischen Funktionen", *Math. Annalen* **14** (1879), 428–471. Reprinted as [Klein 1923, LXXXIV, pp. 90–136]. Translated in this collection.

[Klein 1879b] F. Klein, "Ueber die Transformation elfter Ordnung der elliptischen Modulfunktionen", *Math. Annalen* **15** (1879), 533–555. Reprinted as [Klein 1923, LXXXVI, pp. 140–168].

[Klein 1923] F. Klein, *Gesammelte Mathematische Abhandlungen, 3: Elliptische Funktionen* etc., edited by R. Fricke et al., Springer, Berlin, 1923. Reprinted by Springer, 1973.

[Klein and Fricke 1890–92] F. Klein, *Vorlesungen über die Theorie der elliptischen Modulfunctionen,* ausgearbeitet und vervollständigt von Robert Fricke (2 vol.), Teubner, Leipzig, 1890–92. Reprinted by Johnson Reprint, New York, 1966.

[Levasseur and Stafford 1995] T. Levasseur and J. T. Stafford, "Invariant differential operators and an homomorphism of Harish-Chandra", *J. Amer. Math. Soc.* **8**:2 (1995), 365–372.

[Luna 1975] D. Luna, "Adhérences d'orbite et invariants", *Invent. Math.* **29**:3 (1975), 231–238.

[Wallach 1993] N. R. Wallach, "Invariant differential operators on a reductive Lie algebra and Weyl group representations", *J. Amer. Math. Soc.* **6**:4 (1993), 779–816.

[Weil 1964] A. Weil, "Sur certains groupes d'opérateurs unitaires", *Acta Math.* **111** (1964), 143–211. Reprinted as pp. 1–69 in his *Collected Papers*, vol. III, Springer, New York, 1979.

Allan Adler
P.O.Box 1043
Bowling Green, KY 42102-1043
United States
adler@hera.wku.edu

The Eightfold Way
MSRI Publications
Volume **35**, 1998

Hirzebruch's Curves F_1, F_2, F_4, F_{14}, F_{28} for $\mathbb{Q}(\sqrt{7})$

ALLAN ADLER

Abstract. We give a detailed proof of Hirzebruch's remarkable result that the symmetric Hilbert modular surface of level $\sqrt{7}$ for $\mathbb{Q}(\sqrt{7})$ is $\mathrm{PSL}_2(\mathbb{F}_7)$-equivariantly isomorphic to the complex projective plane. We identify the curves F_1, F_2, F_4, F_{28} explicitly as plane curves defined by invariants of degrees 4,12,18,21 for a three-dimensional representation of $\mathrm{PSL}_2(\mathbb{F}_7)$, and we explain their geometry. For example, F_1 is the Klein curve, F_{12} is the Steinerian of the Klein curve and F_{18} is essentially the Caylean of the Klein curve. The curves F_{12} and F_{18} are birationally equivalent to the Hessian of the Klein curve, which was shown to be defined by a cocompact arithmetic group by Fricke; Hirzebruch's theory gives another uniformization using subgroups of $\mathrm{SL}_2(\mathbb{Z})$. We compute the group of invariant line bundles on the Hessian and offer the Hessian as a challenge to extending Doglachev's recent work on the invariant vector bundles on modular curves to the case of triangle groups $\{p, q, r\}$ in which p, q, r are not pairwise relatively prime. The curve F_{14} maps to the 21-point orbit in $\mathbb{P}^2$. Using our explicit identification of $F_1, F_2, F_4, F_{14}, F_{28}$, we are able to complete Hirzebruch's identification of the nonsymmetric Hilbert modular surface.

Contents

Introduction

Hirzebruch [1977] proved the remarkable result that the complex projective plane $\mathbb{P}^2$ is a minimal model of the symmetric Hilbert modular surface of level $\sqrt{7}$ for the extended Hilbert modular group of $\mathbb{Q}(\sqrt{7})$. Furthermore, the identification is equivariant for natural actions of $\mathrm{PSL}_2(\mathbb{F}_7)$ on the two surfaces. This result also enabled him to identify the (nonsymmetric) Hilbert modular surface associated to this group: namely, it is obtained from $\mathbb{P}^2$ by a certain sequence of blowings up and then passing to a two-sheeted covering branched along a certain curve on the resulting surface.

The curve in question and the sequence of blowings up are given in terms of some $\mathrm{PSL}_2(\mathbb{F}_7)$ orbits and certain $\mathrm{PSL}_2(\mathbb{F}_7)$ invariant curves. These curves are defined from the point of view of the theory of Hilbert modular surfaces as the images of the so-called modular curves F_N in $\mathbb{P}^2$ for $N = 1, 2, 4$. However, the construction of the Hilbert modular surface starting with $\mathbb{P}^2$ fails to be completely explicit since the invariant curves involved in the construction were not completely identified in terms of the geometry of $\mathbb{P}^2$ and $\mathrm{PSL}_2(\mathbb{F}_7)$ alone. The identification of modular curves F_N on a Hilbert modular surface is a matter of independent interest.

The purpose of the present article is twofold. First, we show how to identify the images of the curves F_N in $\mathbb{P}^2$ for $N = 1, 2, 4, 14, 28$. Second, since the details of Hirzebruch's theorem were never published, we provide those details here as a public service. In this latter endeavor, I relied on some unpublished notes and private communications [Hirzebruch 1995; 1979]. Hirzebruch also encouraged me to provide details of his unpublished determination of the curves F_{14} and F_{28}, which were also discussed in [Hirzebruch 1995].

In 1979 or so, I learned of this work of Hirzebruch from a letter of Serre and began looking at [Hirzebruch 1977]. From the sketches there, I was motivated to study the images in $\mathbb{P}^2$ of the curves F_1, F_2 and F_4 knowing only that they were $\mathrm{PSL}_2(\mathbb{F}_7)$ invariant plane curves of degrees 4, 12 and 18 respectively and that their genera were 3, 10 and 10 respectively. The first curve, as Hirzebruch already pointed out, must be Klein's curve

$$x^3y + y^3z + z^3x = 0,$$

but the nature or identity of the other two is not so easy to determine.

I was able to identify the curve F_2 explicitly as the Steinerian of the Klein curve and to write down its equation. (See the beginning of Section 15 for the definition of the Steinerian.) I also showed that there were essentially two possibilities for the singularities of the curve F_4: either it had double points on the 21-point orbit and the 42-point orbit or it had quadruple points on the 21-point orbit. Through a careless error, I incorrectly concluded that the latter

1991 *Mathematics Subject Classification.* 14G35, 14H40, 14H45, 14H60, 14J26, 14J50.

possibility did not occur and only noticed the error after sending the manuscript [Adler 1979] to Hirzebruch. When I pointed out my error to Hirzebruch, he replied that [Berzolari 1903–15] has a footnote describing a way to get a covariant of degree 18 and genus 10 with 21 quadruple points and suggested that this might be the way to obtain F_4 from the Klein curve, since he knew from his study of the Hilbert modular surface that the image of the curve F_4 in $\mathbb{P}^2$ must have 21 quadruple points.

There the matter stood until recently when I again took up the task of completing this article and of providing details of Hirzebruch's results. This effort was partly motivated by my conviction that his striking result was perfect for inclusion in the present volume on the Klein curve.

The effort of providing details given my limited experience with Hilbert modular surfaces turned out to be considerable, partly because of the difficulty of reading and using the relevant literature. Accordingly, I am preparing a set of lecture notes [Adler $\geq$ 1998], which I hope will make it easier for others to gain access to this beautiful subject. The notes will also contain a much more detailed and general examination of the Hilbert modular surface of level $\sqrt{7}$ for $\mathbb{Q}(\sqrt{7})$ than is possible in this article.

The first part of the article is devoted to a proof of Hirzebruch's published results, including the fact that the projective plane is a minimal model of the symmetric Hilbert modular surface of level $\sqrt{7}$ for $\mathbb{Q}(\sqrt{7})$. The second part is devoted to a determination of the curves F_1, F_2, F_4. Some of the computations were carried out using the algebra package REDUCE 3.4 on a personal computer. There are also results regarding the nonsingular models of the curves F_2 and F_4. It turns out that they are isomorphic to the Hessian of the Klein curve and that this is the unique curve of genus 10 with $\mathrm{PSL}_2(\mathbb{F}_7)$ acting on it. We also study the group of invariant line bundles on this curve.

As this article was nearing publication, I ran across [Fricke 1893a] and learned that some of these results on the various models of the Hessian of the Klein curve were anticipated by Fricke a century ago. In particular, Fricke knew that the Hessian of the Klein curve is characterized by its genus and automorphism group. He also wrote down the equations of the unique invariant curve of degree 12 and genus 10 and considered the pencil of all invariant curves of degree 12. Thus, in effect, he found the equations of Hirzebruch's curve F_2. He also considered the pencil of all invariant curves of degree 14 and, as we do, the net of all invariant curves of degree 18 but stopped short of the extensive computations of that net which we have carried out. It will be pleasant to learn more from our late colleague about his old work in this field which for us is so new.

1. Some Hilbert Modular Surfaces for $\mathbb{Q}(\sqrt{7})$

Throughout this article, k will denote the real quadratic field $\mathbb{Q}(\sqrt{7})$ of discriminant $D = 28$ and $\mathcal{O}_k$ will denote the ring of integers of k. The conjugate over $\mathbb{Q}$ of an element x of k or of a matrix M with entries in k will be denoted x' and M' respectively. Denote by $\widehat{\Gamma}$ the group of all 2×2 matrices

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with entries in $\mathcal{O}_k$ such that the determinant of γ is a totally positive unit of k. Denote by $\widehat{\Gamma}(\sqrt{7})$ the subgroup of $\widehat{\Gamma}$ consisting of matrices which are congruent to the identity matrix modulo $\sqrt{7}$. We will refer to $\widehat{\Gamma}$ as the *extended Hilbert modular group*. (The adjective "extended" refers to the fact that $\widehat{\Gamma}$ contains the usual Hilbert modular group $\mathrm{SL}_2(\mathcal{O}_k)$.) It is known [van der Geer 1988, § I.4, pp. 11–14] that $\widehat{\Gamma}$ is a maximal discrete subgroup of $\mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$. We will refer to $\widehat{\Gamma}(\sqrt{7})$ as the *congruence subgroup of level* $\sqrt{7}$ of $\widehat{\Gamma}$. If Γ is a subgroup of $GL_2^+(k)$ (the group of 2×2 matrices with entries in k and totally positive determinant) commensurable with $\mathrm{SL}_2(\mathcal{O}_k)$, we will call Γ a *group of Hilbert modular type*.

A group Γ of Hilbert modular type acts on the product $\mathbb{P}^1(\mathbb{C}) \times \mathbb{P}^1(\mathbb{C})$ of two copies of the projective line $\mathbb{P}^1(\mathbb{C})$ by the rule

$$\gamma\cdot(z_1, z_2) = \left(\frac{az_1 + b}{cz_1 + d}, \frac{a'z_2 + b'}{c'z_2 + d'}\right).$$

We denote by $\mathcal{H}$ the upper half plane in $\mathbb{C}$ and by $\mathcal{H}^2$ the product of two copies of $\mathcal{H}$. If Γ is of Hilbert modular type, we denote the orbit space for Γ acting on $\mathcal{H}^2$ by $\Gamma\backslash\mathcal{H}^2$ and we will refer to any complex analytic surface bimeromorphic to $\Gamma\backslash\mathcal{H}^2$ as a *surface of Hilbert modular type*.

The group $\widehat{\Gamma}(\sqrt{7})$ acts without fixed points on $\mathcal{H}^2$ and the orbit space is a complex manifold. The group $\widehat{\Gamma}$ has some isolated fixed points arising from elliptic elements and its orbit space is a complex orbifold but is not a complex manifold due to the singularities arising from the elliptic fixed points. (For the notion of an orbifold see [Satake 1956; Weil 1962], where the terminology "V-manifold" is used.)

One can prove that both orbit spaces are isomorphic to quasiprojective algebraic varieties. The proof depends on showing that both spaces have natural compactifications which are isomorphic to projective varieties. We will not present the proof of this result, but the compactifications themselves [Satake 1960; Baily and Borel 1966] are of interest to us. They are defined in [van der Geer 1988, § I.4], and we refer the reader there for details. The compactification and its structure of normal complex analytic variety were constructed in a general setting by Satake [1960]. That complex analytic variety is called the *Satake compactification*. Baily and Borel [1966] proved that the Satake compactification is isomorphic to a projectively normal (hence normal) projective variety.

That variety is called the *Baily–Borel compactification*, but since we don't want Satake's role to be forgotten, we will refer to it as the *SBB compactification* throughout this article.

We will denote the SBB compactifications of $\widehat{\Gamma}\backslash\mathcal{H}^2$ and $\widehat{\Gamma}(\sqrt{7})\backslash\mathcal{H}^2$ respectively by $\widehat{X}$ and $\widehat{X}(\sqrt{7})$. Write Γ to denote either $\widehat{\Gamma}$ or $\widehat{\Gamma}(\sqrt{7})$ and X to denote the SBB compactification of $\Gamma\backslash\mathcal{H}^2$. Then the complement of $\Gamma\backslash\mathcal{H}^2$ in X is a finite set, namely the orbit space for Γ acting on the projective line $\mathbb{P}^1(k)$ over k. This projective line sits in $\mathbb{P}^1(\mathbb{C})\times\mathbb{P}^1(\mathbb{C})$ via the embedding $\xi\mapsto(\xi,\xi')$. The orbits for Γ on $\mathbb{P}^1(k)$ will be called the *cusps* of Γ or of X. The point (∞,∞) is a cusp. It and the points corresponding to it in surfaces of Hilbert modular type will be denoted ∞.

Since the class number of k is 1, the number of cusps of $\widehat{\Gamma}$ is 1 [van der Geer 1988, Prop. I.1.1, p. 6], and the number of cusps of $\widehat{\Gamma}(\sqrt{7})$ is $(7^2-1)/2=24$.

The natural mapping of $\widehat{X}(\sqrt{7})$ onto $\widehat{X}$ has degree equal to the order of $\mathrm{PSL}_2(\mathbb{F}_7)$, which is 168. For the rest of this article, we will denote the group $\mathrm{PSL}_2(\mathbb{F}_7)$ by G.

2. Some Congruence Subgroups of Unit Groups of Orders in Quaternion Algebras

Recall that $k=\mathbb{Q}(\sqrt{7})$, $D=28$ is the discriminant of k and $\mathcal{O}_k$ is its ring of integers. For the rest of this article, $\eta=3-\sqrt{7}$ is an element of norm 2 and $\varepsilon=\eta/\eta'=8-3\sqrt{7}$ is a fundamental unit of k.

Definition 2.1. By a *skew-hermitian matrix* we will mean a matrix of the form

$$B=\begin{pmatrix} a\sqrt{D} & \lambda \\ -\lambda' & b\sqrt{D}\end{pmatrix},$$

where a,b are rational numbers and λ is an element of k. We will say such a matrix is *integral* if a,b are rational integers and λ is an integer of k. We will say an integral skew-hermitian matrix B is *primitive* if there is no rational integer $n>1$ such that $\frac{1}{n}B$ is also integral.

The proofs of the next two lemmas are left to the reader.

Lemma 2.2. *If ν is a nonnegative integer, the skew-hermitian matrix B_ν given by*

$$\begin{pmatrix} 0 & \eta^\nu \\ -\eta'^\nu & \sqrt{D}\end{pmatrix}$$

is primitive and has determinant 2^ν.

Lemma 2.3. *If ν is a nonnegative integer, the matrix B_ν given by*

$$\begin{pmatrix} 2^\nu\sqrt{D} & 0 \\ 0 & \sqrt{D}\end{pmatrix}$$

is a primitive skew-hermitian matrix of determinant $2^\nu D$.

Lemma 2.4. *There is no integral skew-hermitian form B with determinant 7 but there is one of determinant 14.*

Proof. The skew-hermitian matrix B given by

$$\begin{pmatrix} \sqrt{D} & 7-\sqrt{7} \\ -7-\sqrt{7} & -\sqrt{D} \end{pmatrix} \tag{2.5}$$

has determinant 14. We have $\mathcal{O}_k = \mathbb{Z}[\sqrt{7}]$. The general integral skew-hermitian matrix

$$\begin{pmatrix} a\sqrt{D} & \lambda \\ -\lambda' & b\sqrt{D} \end{pmatrix}$$

has determinant $28ab + \nu(\lambda)$, where ν denotes the norm of the quadratic field k. If this determinant equals 7 then 7 divides the norm $\lambda\lambda'$ of λ, whence $\sqrt{7}$ divides λ. Writing

$$\lambda = (c + d\sqrt{7})\sqrt{7},$$

we have

$$4ab - c^2 + 7d^2 = 1,$$

where a, b, c, d are rational integers. Modulo 4 this becomes

$$-c^2 - d^2 \equiv 1,$$

which clearly has no solution. $\square$

Proposition 2.6. *Let B be a primitive integral skew-hermitian matrix with entries in k. Denote by Q_B the set of all 2×2 matrices*

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with entries in k such that

$${}^tM'B = BM^*, \tag{2.7}$$

where

$$M^* = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Then Q_B is an indefinite quaternion algebra over $\mathbb{Q}$ generated by elements i, j such that

$$i^2 = D, \quad j^2 = -N/D, \quad ij = -ji,$$

where N is the determinant of B. The intersection of Q_B with the ring of 2×2 matrices with entries in $\mathcal{O}_k$ is an order, denoted $\mathcal{O}_B$, of Q_B. The discriminant of the order $\mathcal{O}_B$ is N^2.

(The discriminant of the order $\mathcal{O}_B$ is defined to be the index of $\mathcal{O}_B$ in its dual lattice in Q_B with respect to the pairing defined by the reduced trace of Q_B. This is not the same as the discriminant of the quaternion algebra Q_B, which is the product of the primes q such that Q_B is ramified at q. The discriminant of a maximal order of Q_B is the square of the discriminant of Q_B.)

PROOF. See [van der Geer 1988, Prop. V.1.5, p. 90]. □

LEMMA 2.8. *If B is as in Lemma 2.2 then the quaternion algebra Q_B is a matrix algebra over $\mathbb{Q}$. If B is as in Lemma 2.3 or Equation 2.5 then Q_B is a division algebra with discriminant 14.*

PROOF. Since k has discriminant $D = 28$, it follows from Lemma 2.6 that the reduced norm for Q_B is of the form

$$w^2 - 28x^2 + \tfrac{1}{7}\cdot 2^{\nu-2}y^2 - 2^\nu z^2$$

with $\nu \geq 0$ or, after replacing x by $x/2$ and y by $14y$,

$$\xi\xi' - 2^\nu \zeta\zeta',$$

where $\xi = w + x\sqrt{7}$ and $\zeta = z + y\sqrt{7}$. Since $\eta\eta' = 2$ and the norm from k to $\mathbb{Q}$ is multiplicative, the form is equivalent to

$$\xi\xi' - \zeta\zeta',$$

which obviously represents 0. This proves the first assertion.

If B is B_ν as in Lemma 2.3 or B as in Equation 2.5 then by Lemma 2.7, the norm form is

$$w^2 - 28x^2 + 2^{\nu-2}y^2 - 2^\nu \cdot 7z^2$$

with $\nu \geq -1$, which is equivalent to

$$\xi\xi' + \zeta\zeta'$$

with $\xi = w + x\sqrt{7}$ and $\zeta = y + z\sqrt{7}$, by a reduction similar to the preceding one, and hence to

$$w^2 - 7x^2 + y^2 - 7z^2.$$

It is easy to see that this form cannot represent zero nontrivially. Indeed, one can assume that w, x, y, z are all integers without common factor. Reducing modulo 7, one concludes that 7 must divide w and y. Replacing w by $7w$, y by $7y$, dividing through by 7 and again reducing modulo 7, one concludes that x, z are divisible by 7, which contradicts the assumption that w, x, y, z are integers without common factor. □

LEMMA 2.9. *Let B be as in Lemma 2.8. With the notation of Lemma 2.6, denote by $\mathcal{O}_B^1$ the group of units of $\mathcal{O}_B$ of reduced norm 1. Denote by $\mathcal{O}_B^1(\sqrt{7})$ the intersection of the group of $\mathcal{O}_B^1$ consisting of elements which are congruent to 1 with the group $\widehat{\Gamma}(\sqrt{7})$. If B is B_ν as in Lemma 2.2 then the factor group $\mathcal{O}_B^1/\{\pm 1\}\mathcal{O}_B^1(\sqrt{7})$ is isomorphic to G. If B is B_ν as in Lemma 2.3 or B as in Equation 2.5, then the factor group is a cyclic group of order 4.*

Proof. By the approximation theorem, the factor group $\mathcal{O}_B^1/\mathcal{O}_B^1(\sqrt{7})$ is the product of the analogous groups for the q-adic completions of the algebra Q_B, as q runs over all primes. In particular, it must contain the analogous group for the 7-adic completion. On the other hand, the factor group $\mathcal{O}_B^1/\mathcal{O}_B^1(\sqrt{7})$ clearly injects into the factor group $\widehat{\Gamma}/\widehat{\Gamma}(\sqrt{7})$, which is isomorphic to $\mathrm{SL}_2(\mathbb{F}_7)$. Now suppose we are in the case of $B = B_\nu$ as in Lemma 2.2. Then Q_B is a matrix algebra isomorphic to $M_2(\mathbb{Q})$. According to Lemma 2.2, the discriminant of the order $\mathcal{O}_B$ is a power of 2, hence its 7-adic completion is a maximal order of $M_2(\mathbb{Q}_7)$ and the 7-adic completion of $\mathcal{O}_B^1$ is $\mathrm{SL}_2(\mathbb{Z}_7)$. Similarly, since $i^2 = D = 28$, the 7-adic completion of the congruence subgroup $\mathcal{O}_B^1(\sqrt{7})$ is the congruence subgroup of level 7 of $\mathrm{SL}_2(\mathbb{Z}_7)$ and the factor group is $\mathrm{SL}_2(\mathbb{F}_7)$. Therefore, the contribution at 7 to the factor group $\mathcal{O}_B^1/\mathcal{O}_B^1(\sqrt{7})$ is $\mathrm{SL}_2(\mathbb{F}_7)$ and to the factor group $\mathcal{O}_B^1/\{\pm 1\}\mathcal{O}_B^1(\sqrt{7})$ is G, which implies that $\mathcal{O}_B^1/\{\pm 1\}\mathcal{O}_B^1(\sqrt{7})$ must be all of G in this case. If B is B_ν as in Lemma 2.3 or B as in Equation 2.5, the quaternion algebra Q_B is ramified at 7 by Lemma 2.6. The discriminant of the order $\mathcal{O}_B$ is a power of 2 times the square of the discriminant of the algebra Q_B. Therefore, the 7-adic completion of $\mathcal{O}_B$ is a maximal order of the 7-adic completion of Q_B. The 7-adic completion of $\mathcal{O}_B^1$ is then the group of units of reduced norm 1 in that maximal order and the 7-adic completion of $\mathcal{O}_B^1(\sqrt{7})$ is the subgroup of elements congruent to 1 modulo the maximal ideal of that maximal order. The maximal order modulo its maximal ideal is isomorphic to the field $\mathbb{F}_{49}$ with $7^2 = 49$ elements [Weil 1974, Prop. I.4.5, pp. 20-21]. The factor group $\mathcal{O}_B^1/\mathcal{O}_B^1(\sqrt{7})$ is therefore isomorphic to the group of elements of norm 1 in $\mathbb{F}_{49}$. Since the norm map is a surjective homomorphism from $\mathbb{F}_{49}^\times$ to $\mathbb{F}_7^\times$, the elements of norm 1 form a group of order $(49-1)/(7-1) = 8$, in fact a cyclic group, and that cyclic group is therefore contained in the factor group $\mathcal{O}_B^1/\mathcal{O}_B^1(\sqrt{7})$. When we pass to the factor group $\mathcal{O}_B^1/\{\pm 1\}\mathcal{O}_B^1(\sqrt{7})$, the contribution from the 7-adic completion is therefore a cyclic group of order 4. This cyclic group is a maximal commutative subgroup of the group G. In particular, it is its own centralizer in G. Since the group $\mathcal{O}_B^1/\mathcal{O}_B^1(\sqrt{7})$ is the product of this cyclic group with the analogous groups at primes other than 7, the latter must commute with the cyclic group. It follows that the groups coming from the other primes must be trivial. □

If B is a primitive skew-Hermtitian matrix with positive determinant, we denote by $\widehat{\Gamma}_B$ the subgroup of $\widehat{\Gamma}$ consisting of similitudes of B, i.e.,

$$\widehat{\Gamma}_B = \{M \in \widehat{\Gamma} \mid (\exists \xi \in k)\, {}^tM'BM = \xi B\}. \tag{2.10}$$

Using the notation of equation (2.7), for invertible matrices M we have $M^* = M^{-1} \cdot \det(M)$. Replacing ξ by $\xi \cdot \det(M)$ in (2.10), we can therefore rewrite (2.10) as

$$\widehat{\Gamma}_B = \{M \in \widehat{\Gamma} \mid (\exists \xi \in k)\, {}^tM'B = \xi B M^*\}, \tag{2.11}$$

which is a natural generalization of (2.7). It is easy to verify that $\widehat{\Gamma}_B$ is a subgroup of $\widehat{\Gamma}$.

LEMMA 2.12. *Let B be as in Lemma* 2.3 *or equation* (2.5). *Then the image of $\widehat{\Gamma}_B$ in G is a Sylow* 2*-subgroup of G and therefore has order* 8.

PROOF. We will refer to B in Lemma 2.3 as Case 1 and B as in equation (2.5) as Case 2. Let $C = \frac{1}{\sqrt{7}}B$. Then we have

$$C = \begin{pmatrix} 2^{\nu+1} & 0 \\ 0 & 2 \end{pmatrix}$$

if B in Case 1 and

$$C = \begin{pmatrix} 2 & -1+\sqrt{7} \\ -1-\sqrt{7} & -2 \end{pmatrix}$$

in Case 2. Clearly, in the definition of $\widehat{\Gamma}_B$ we can replace B by C. Let $\overline{C}$ denote the reduction of C modulo $\sqrt{7}$. Since $x' \equiv x$ modulo $\sqrt{7}$ for all x in $\mathcal{O}_k$, the matrix $\overline{C}$ will be a symmetric matrix with entries in $\mathbb{F}_7$. Explicitly,

$$\overline{C} = \begin{pmatrix} 2^{\nu+1} & 0 \\ 0 & 2 \end{pmatrix}$$

in Case 1 and

$$\overline{C} = \begin{pmatrix} 2 & -1 \\ -1 & -2 \end{pmatrix}$$

in Case 2. Therefore, the image of $\widehat{\Gamma}_B$ in $\widehat{\Gamma}/\widehat{\Gamma}(\sqrt{7}) = \mathrm{SL}_2(\mathbb{F}_7)$ lies in the group of similitudes of determinant 1 of the quadratic form determined by $\overline{C}$. That quadratic form is $2^{\nu+1}x^2 + 2y^2$ in Case 1 and is $2x^2 - 2xy - 2y^2$ in Case 2. Using the fact that 2 is a square in $\mathbb{F}_7$ it is easy to verify that in either case the quadratic form is anisotropic and therefore equivalent, up to scalar multiple, to the quadratic form x^2+y^2 whose group of orthogonal similitudes of determinant 1 is easily seen to be of order 16 and is a Sylow 2-subgroup of $\mathrm{SL}_2(\mathbb{F}_7)$. Its image in $\mathrm{PSL}_2(\mathbb{F}_7)$ is therefore a Sylow 2-subgroup of of $\mathrm{PSL}_2(\mathbb{F}_7)$. It remains to show that the group $\widehat{\Gamma}_B$ maps onto this subgroup. From Lemma 2.9, we know that the subgroup $\mathcal{O}_B^1$ maps onto a cyclic subgroup of $\mathrm{PSL}_2(\mathbb{F}_7)$ of order 4. Since the elements of $\mathcal{O}_B^1$ consist of elements of $\widehat{\Gamma}_B$ which satisfy (2.10) with $\xi \equiv 1$ modulo $\sqrt{7}$, it is enough to show that $\widehat{\Gamma}_B$ contains an element which satisfies (2.10) with $\xi = -1$. But this is in fact a special case of [Hausmann 1980, Cor. 2.9, pp. 14–15]. □

3. Modular Curves on a Hilbert Modular Surface

If B is a primitive integral skew-hermitian matrix over k, we will denote by $\mathcal{H}_B$ the locus in $\mathcal{H}^2$ consisting of all $z = (z_1, z_2)$ in $\mathcal{H}^2$ such that

$$(z_2\ 1)\ B \begin{pmatrix} z_1 \\ 1 \end{pmatrix} = 0. \tag{3.1}$$

If S is a complex surface with a mapping ϕ (usually understood from the context) of $\mathcal{H}^2$ into S, then the closure of the image $\phi(\mathcal{H}_B)$ of $\mathcal{H}_B$ in S under ϕ will be denoted $F_B(S)$. For example, $F_B(\mathcal{H}^2) = \mathcal{H}_B$.

Let N be a positive integer. The union of all $F_B(S)$ as B runs over all primitive integral skew-hermitian forms of determinant N will be denoted $F_N(S)$. The union of all F_{N/d^2}, where d runs over all positive integers such that d^2 divides N, will be denoted $T_N(S)$. It is the same to say that one obtains T_N by omitting the primitivity condition from the definition of F_N.

In most of the cases of interest to us (though not all, since some or all of the curves may collapse to points), the loci $F_B(S)$, $F_N(S)$ and $T_N(S)$ will be complex analytic curves on the surface S. We will be particularly interested in cases where S is an algebraic surface and in that case, these loci will actually be algebraic.

We need to know the number of components of the curves $F_N(S)$ in certain cases. For this it is helpful to know the subgroup of $\widehat{\Gamma}$ leaving $\mathcal{H}_B$ invariant. It turns out to be a group with which we are already familiar.

LEMMA 3.2. *The subgroup of* $\widehat{\Gamma}$ *leaving* $F_B(\mathcal{H}^2)$ *invariant is* $\widehat{\Gamma}_B$ (see Equation 2.10).

PROOF. Suppose $\gamma \in \widehat{\Gamma}$. Then γ^{-1} maps the locus of of all $z = (z_1, z_2)$ in $\mathcal{H}^2$ such that

$$0 = (z_2 \ 1) \ B \begin{pmatrix} z_1 \\ 1 \end{pmatrix}$$

to the locus of all z such that

$$0 = (z_2 \ 1) \ {}^t\gamma' B \gamma \begin{pmatrix} z_1 \\ 1 \end{pmatrix}.$$

Since $F_B(\mathcal{H}^2)$, when nonempty, determines B up to a scalar factor, the lemma follows at once. $\square$

LEMMA 3.3. *Let* $\phi : \widehat{\Gamma}(\sqrt{7})\backslash\mathcal{H}^2 \to S$ *be a holomorphic mapping with dense image. Then the curve* $F_7(S)$ *is empty. Let* $r \geq 1$ *and* $N = 2^r \cdot 7$. *Then the curve* $F_N(\widehat{X}(\mathcal{O}_k))$ *has exactly one component. For* $r \geq 0$ *and* $N = 2^r$, *the curve* $F_N(\widehat{X}(\mathcal{O}_k))$ *has just* 1 *component for* $r \leq 3$ *and exactly* 2 *components for* $r \geq 4$.

PROOF. The assertion about F_7 follows at once from the definition and from Lemma 2.4. By Lemmas 2.2, 2.3, 2.4, there do exist primitive skew-hermitian forms over k with determinant $N = 2^r \cdot 7$ with $r \geq 1$ and determinant $N = 2^r$ with $r \geq 1$. Therefore the loci $F_N(S)$ are nonempty for such N. The number of components is now determined using the table in [Hausmann 1980, p. 20] (cf. [van der Geer 1988, § V.3, pp. 93–100]). $\square$

COROLLARY 3.4. *The number of components of* $F_N(\widehat{X}(\sqrt{7}))$, *with* $N = 1, 2, 4$ *is* 1. *The number of components of* $F_N(\widehat{X}(\sqrt{7}))$ *with* $N = 2^r \cdot 7$, $r \geq 1$, *is* 21.

PROOF. By Lemma 3.3, $F_N(\widehat{X})$ has only one component. Therefore, by Lemma 3.2 the set of components of $F_N(\mathcal{H}^2)$ can be identified $\widehat{\Gamma}$-equivariantly with $\widehat{\Gamma}/\widehat{\Gamma}_B$, where B is as in Lemma 2.2, Lemma 2.3 or equation (2.5) according to the value of N. It follows that the set of components of $F_N(\widehat{X}(\sqrt{7}))$ can be identified with the double coset space $\{\pm 1\}\widehat{\Gamma}(\sqrt{7})\backslash\widehat{\Gamma}/\widehat{\Gamma}_B$. Since $\{\pm 1\}\widehat{\Gamma}(\sqrt{7})$ is a normal subgroup of $\widehat{\Gamma}$ with quotient G, it follows that the number of components of $F_N(\widehat{X}(\sqrt{7}))$ is the index in G of the image of $\widehat{\Gamma}_B$ in G. By Lemma 2.9, the image of $\widehat{\Gamma}_B$ in G is all of G if $N = 1, 2, 4$, which proves the first assertion of the lemma. If $N = 2^r \cdot 7$ with $r \geq 1$, then by Lemma 2.12, the image of $\widehat{\Gamma}_B$ has order 8, so $F_N(\widehat{X}(\sqrt{7}))$ has $168/8 = 21$ components. □

LEMMA 3.5. *Let*

$$B = \begin{pmatrix} a\sqrt{D} & \lambda \\ -\lambda' & b\sqrt{D} \end{pmatrix}$$

be a primitive integral skew-hermitian matrix over k with positive determinant. Then the analytic disc $\mathcal{H}_B$ has ∞ as a limit point if and only if $a = 0$.

PROOF. The equation (3.1) defining $\mathcal{H}_B$ when expanded becomes

$$az_1z_2\sqrt{D} + \lambda z_2 - \lambda' z_1 + b\sqrt{D} = 0. \tag{3.6}$$

Dividing through by z_1z_2 we see that if (z_1, z_2) can approach ∞ along $\mathcal{H}_B$, in the limit we have $a = 0$. Coversely, if $a = 0$ then $\lambda \neq 0$ and z_2 will approach ∞ along with z_1. □

4. Volumes and Genera of Modular Curves on $\widehat{X}(\sqrt{7})$

In the following lemma, the characters χ_{-7} and χ_{-4} are Dirichlet characters modulo 7 and modulo 4 respectively and are defined as follows. Letting t denote either 7 or 4, let ζ_t denote a primitive t-th root of unity and $\mathbb{Q}(\zeta_t)$ the cyclotomic field of t-th roots of unity. The value of $\chi_{-t}(n)$ is 0 if t divides n. To compute $\chi_{-t}(n)$ for n prime to t, consider the automorphism σ_n of $\mathbb{Q}(\zeta_t)$ given by $\zeta_t \mapsto \zeta_t^n$. Then the value of $\chi_{-t}(n)$ is 1 or -1 according to whether the automorphism σ_n does or does not induce a trivial automorphism of the quadratic subfield $\mathbb{Q}(\sqrt{-t})$ of $\mathbb{Q}(\zeta_t)$. The Legendre symbol $\left(\frac{D}{q}\right)$ and the Hilbert symbol $\left(\frac{N,D}{q}\right)$ appear in the definition of α_q.

LEMMA 4.1. *The volume of $F_N(\widehat{X})$ with respect to the volume form*

$$-\frac{1}{2\pi}\frac{dx \wedge dy}{y^2}$$

is

$$\operatorname{vol}(F_N(\widehat{X})) = -\frac{1}{24}N(1 + \chi_{-p}(N))(1 + \chi_{-4}(N))\prod_{q|N}\alpha_q,$$

where α_q is defined for primes q dividing N by

$$\alpha_q = \begin{cases} 1 + \left(\dfrac{D}{q}\right)/q & \text{if } q\nmid D \\ 1 + \left(\dfrac{N,D}{q}\right)/q & \text{if } q|D \text{ and } q||N \\ 1 - \dfrac{1}{q^2} & \text{if } q|D \text{ and } q^2|N. \end{cases}$$

Proof. In [Hausmann 1980, p. 49] (cf. [van der Geer 1988, § V.5, pp. 101–102]) one finds a volume formula which specializes to the one given above except for having the fraction on the right-hand side equal to $-1/12$ instead of $-1/24$. However, the formula of Hausmann was for the volume of the curve $F_N(X)$, where X is a compactification of $\mathrm{SL}_2(\mathcal{O}_k)\backslash\mathcal{H}^2$. Therefore, $F_N(X)$ is a two-sheeted cover of the curve $F_N(\widehat{X})$, because $\mathrm{PSL}_2(\mathcal{O}_k)$ is isomorphic to a subgroup of index 2 in $\widehat{\Gamma}/\mathcal{O}_k^\times$. Therefore our curve has half the volume given by Hausmann's formula. $\square$

Corollary 4.2. *Let $k = \mathbb{Q}(\sqrt{7})$ and let J be the two sided ideal $(\sqrt{7}) = \sqrt{7}\mathcal{O}_k$ of $\mathcal{O}_k$. Then the volumes of the curves $F_N(\widehat{X}(\sqrt{7}))$ for $N = 1, 2, 4, 7, 14, 2^r\cdot 7$ are given by*:

N	1	2	$2^r,\ r\geq 2$	7	14	$2^r\cdot 7,\ r\geq 2$
$\mathrm{vol}(F_N(\widehat{X}(\mathcal{O}_k)))$	-28	-42	$-2^{r-1}\cdot 21$	0	-42	$-2^{r-1}\cdot 63$

Proof. Since $\widehat{X}(\sqrt{7})$ is a 168-sheeted branched cover of $\widehat{X}$, with branching only at a finite set of points, the volume of the curve $F_N(\widehat{X}(\sqrt{7}))$ is 168 times the volume of $F_N(\widehat{X})$. The latter is computed directly from Lemma 4.1. We leave the arithmetic to the reader. $\square$

Corollary 4.3. *For $N = 2^r\cdot 7$, with $r\geq 1$, the curve $F_N(\widehat{X}(\sqrt{7}))$ consists of exactly 21 nonsingular irreducible components. Each component of $F_{14}(\widehat{X}(\sqrt{7}))$ has volume -2 and genus 2. If $r > 1$, each component of $F_N(\widehat{X}(\sqrt{7}))$ has volume $-2^{r-1}\cdot 3$ and genus $1 + 2^{r-2}\cdot 3$.*

Proof. The volume of $F_N(\widehat{X}(\sqrt{7}))$ for $N = 2^r\cdot 7$, $r\geq 1$, is given in the preceding corollary. By Corollary 3.4 the number of components of $F_N(\widehat{X}(\sqrt{7}))$, for such N is 21. Hence, each component of $F_N(\widehat{X}(\sqrt{7}))$ has volume -2 if $r = 1$ and $-2^{r-1}\cdot 3$ if $r > 1$. Since these volumes coincide with the Euler numbers of the components, we conclude that the genus of each component of $F_N(\widehat{X}(\sqrt{7}))$ is 2 if $r = 1$ and is $1 + 2^{r-2}\cdot 3$ if $r > 1$. $\square$

Remark 4.4. In Lemma 7.15, we will prove that the components of $F_{14}(\widehat{X}(\sqrt{7}))$ and $F_{28}(\widehat{X}(\sqrt{7}))$ are hyperelliptic curves.

5. Intersections of Modular Curves on $\widehat{X}(\sqrt{7})$

In this section, we will compute the intersection numbers of certain of the modular curves $F_N(\sqrt{7})$ on the complete normal singular surface $\widehat{X}(\sqrt{7})$. The theory of intersections on complete normal singular surfaces is due to Mumford [1961] (cf. [Fulton 1984, Ex. 7.1.16, p. 125]). We will have more to say about it in Section 7. Our main tool for computing these intersection numbers is a formula due to Hausmann. However, Hausmann's formula is for intersections of modular curves on the normal singular surface $\widehat{X}$, so we need some preliminary results (Lemmas 5.1 and 5.3), which seem to be implicit in the literature on Hilbert modular surfaces. I am indebted to Angelo Vistoli and Torsten Ekedahl for their help in explaining to me how to prove it and in criticizing my earlier drafts of the proofs. The relevant principles will be discussed in greater detail. in [Adler $\geq$ 1998].

LEMMA 5.1. *If $g : X \to Y$ is a surjective morphism of degree n of complete normal surfaces, with X smooth, and if C is a Weil divisor on Y, then there is a unique Weil divisor C' on X with rational coefficients with the following properties*:

(1) *Let Y_0 be the smooth locus of Y, let $X_0 = g^{-1}(Y_0)$ and let $g_0 : X_0 \to Y_0$ be the restriction of g. Let $C_0 = C \cap Y_0$ and $C' = C' \cap X_0$. Then $C'_0 = g_0^* C_0$.*
(2) *If D is any curve component of the complement of X_0 in X then $(C' \cdot D)_X = 0$.*

PROOF. For $n = 1$, this is due to Mumford [1961] and is the key to his intersection product. In the general case, use the Stein factorization of g:

$$X \to Y' \to Y$$

to first pull C back to C'' on Y' and then to use Mumford's result for $n = 1$ for the morphism $X \to Y'$. The condition (2) follows from the fact that the complement of X_0 in X maps to a finite set in Y'. □

REMARK 5.2. We will call C' the *Mumford pullback* of C and denote it $g^M C$. In the proof of Lemma 5.3, we also need to use another notion of pullback. Let $g : X \to Y$ be a surjective quasifinite morphism of complete normal surfaces and let C be a Weil divisor on Y. Let Y_0 be the smooth locus of Y, let $X_0 = f^{-1}(Y_0)$, let $C_0 = C \cap Y_0$ and let $g_0 : X_0 \to Y_0$ be the restriction of g. Then C_0 is a Cartier divisor on Y_0 whose pullback to X_0 is denoted $g^*(C_0)$. The scheme theoretic closure of $g^*(C_0)$ in X is then associated to a Weil divisor on X which we denote $g^*(C)$. We remark that on any surface, a Weil divisor is uniquely determined by its restriction to the complement of a finite set. This notion of pull-back is a special case of the notion of pullback of codimension-1 cycles presented in [Grothendieck 1967, § 21.10] and will be denoted f^*. In the case of Cartier divisors, it coincides with the usual notion of pullback.

LEMMA 5.3. *Let $f : U \to V$ be a finite surjective morphism of complete normal surfaces. Let n be the degree of f. Let C_1, C_2 be curves on V. Denote by $(\cdot\,,\cdot)_U$ and $(\cdot\,,\cdot)_V$ the Mumford intersection products on U and V respectively. Then if C_i' is the pullback to U of the divisor C_i in U, we have*

$$(C_1', C_2')_U = n\cdot(C_1, C_2)_V.$$

PROOF. Suppose that $\pi_1 : Z \to V$ is a desingularization and that $\pi_2 : W \to U$ is a desingularization factoring through the fibre product of f and π_1. Denote by $h : W \to Y$ the natural mapping from W to Z. Then we have $f\pi_1 = \pi_2 h$. I claim that we also have

$$(f\pi_1)^M = \pi_1^M f^* = h^* \pi_2^M.$$

Indeed, our assumptions imply that $W \to U \to V$ is the Stein factorization of $f\pi_1$, so the first equality is an immediate consequence of our construction of $(f\pi_1)^M$. As for the second, we just have to prove that it satisfies the conditions (1) and (2) above. Let C be a Weil divisor on V. Clearly $h^*\pi_2^M C$ satisfies condition (1). As for condition (2), let D be a curve component of the preimage in W of the singular locus of V. We have

$$(h^*\pi_2^M C\cdot D)_W = (\pi_2^M C\cdot h_*(D))_Z.$$

If h maps D to a point, this last expression is already 0. If h maps D to a curve, that curve is mapped by π_2 to a singular point of V and by definition of π_2^M the right side is again 0. We can now prove the lemma. Let C_1, C_2 be Weil divisors on V. Then

$$\begin{aligned}(f^*C_1\cdot f^*C_2)_U &= (\pi_1^M f^*C_1\cdot\pi_1^M f^*C_2)_W\\ &= (h^*\pi_2^M C_1\cdot h^*\pi_2^M C_2)_W\\ &= n(\pi_2^M C_1\cdot\pi_2^M C_2)_Z\\ &= n(C_1\cdot C_2)_V.\end{aligned}$$

Here the penultimate equality follows from the fact that the lemma is already well known for morphisms between smooth varieties. □

As for Hausmann's formula, it is given below. It is derived in [Hausmann 1980, Satz 5.13, p. 102] (cf. [van der Geer 1988, Cor. VI.5.3, p. 144]) in greater generality than we state it here. The notation appearing in the formula, other than $P_\rho(M, N)$, requires more extensive explanation, which will be given after the statement of the formula. We will discuss Hausmann's formula in greater detail in [Adler $\geq$ 1998].

INTERSECTION FORMULA 5.4. *Let k be a real quadratic field of discriminant D. Let M and N be positive integers. Then the intersection number of T_M and T_N*

on $\widehat{X}$ *(in the sense of Mumford's theory) is given by*

$$(T_M \cdot T_N)_{\widehat{X}} = \sum_{n|(M,N)} \left(n\left(H_D\left(\frac{MN}{n^2}\right) + I_D\left(\frac{MN}{n^2}\right)\right) \prod_{\rho|D} \frac{\chi_{D(\rho)}(n) + \chi_{D(\rho)}(P_\rho(M,N)/n)}{2} \right),$$

where the product runs over all rational primes dividing D *and where*

$$P_\rho(M,N) = \rho^{\min(\mu,\nu)},$$

with $\mu = \operatorname{ord}_\rho M$ *and* $\nu = \operatorname{ord}_\rho N$.

We now explain the notation.

(1) The Dirichlet characters χ_{-4} and χ_{-7} were defined at the beginning of Section 4.

(2) In order to explain the notation H_D, we first have to explain the number theoretic functions h, h' and H.

(a) If $\mathcal{A}$ is an order in an imaginary quadratic field K, we denote by $h(\mathcal{A})$ the class number of $\mathcal{A}$. Since an order is determined up to isomorphism by its discriminant, we will also define $h(d)$ to be $h(\mathcal{A})$ if d is the discriminant of $\mathcal{A}$. We also denote by $\mathcal{A}_d$ the unique order of discriminant d. If d is a rational number which is not the discriminant of an order in an imaginary quadratic field, we define $h(d)$ to be 0.

(b) We define the number theoretic function h' as follows. For all rational numbers d, we define $h'(d)$ by

$$h'(d) = \begin{cases} -1/12 & \text{if } d = 0, \\ 1/2 & \text{if } d = -4, \\ 1/3 & \text{if } d = -6, \\ h(d) & \text{otherwise.} \end{cases}$$

(c) We define the Hurwitz–Kronecker class number H by

$$H(d) = \sum_{c^2|d} h'(-d/c^2)$$

for all rational numbers d.

(d) We define the function H_D by

$$H_D(n) = \sum H\left(\frac{4n - x^2}{D}\right),$$

where the summation on the right runs over all rational numbers x.

(3) Let k be a real quadratic field with discriminant D and let $\mathrm{Id}(k)$ denote the group of fractional ideals of k. Then we define the function $f : \mathrm{Id}(k) \to \mathbb{R}$ by

$$f(M) = \frac{1}{\sqrt{D}} \sum min(\lambda, \lambda'),$$

where the summation runs over all totally positive elements λ of k such that $M = \mathcal{O}_k\lambda$. We define the number theoretic function I_D on the positive rational integers by the rule

$$I_D(N) = \sum f(M)$$

where the summation runs over integral ideals k of norm N.

We now provide some values of H_D and of I_D.

COROLLARY 5.5. *We have the following values of H_D when $D = 28$.*

N	1	2	4	8	16	49	98	196	392	784
$H_D(N)$	$-\frac{1}{6}$	0	$-\frac{1}{6}$	0	$-\frac{1}{6}$	$\frac{5}{6}$	2	$\frac{11}{6}$	8	$\frac{71}{6}$

PROOF. Since $h(7) = 1$ and $h(56) = h(84) = 4$ (see any table of class numbers of imaginary quadratic fields), this follows at once from the following lemma, which is used for computing the class numbers of nonmaximal orders. □

LEMMA 5.6. *Let $\mathcal{A}$ be an order in a quadratic field K. Let h be the class number of K and let $\mathcal{O}_K$ be the maximal order of K. Then the class number of $\mathcal{A}$ is given by*

$$h(\mathcal{A}) = h(\mathcal{O}_K)\frac{|(\mathcal{O}_K/f\mathcal{O}_K)^\times|}{|\mathcal{O}_K^\times/\mathcal{A}^\times|\cdot\varphi(f)},$$

where φ is the Euler φ function.

PROOF. See [Borevich and Shafarevich 1966, pp. 152–153]. □

LEMMA 5.7. $I_D(2^r\cdot 7) = 0$ *for all* $r \geq 0$.

PROOF. There is only one ideal of k with norm $2^r\cdot 7$, namely the one generated by $\eta^r\sqrt{7}$. Since this element has negative norm and since k has no unit of negative norm, the ideal is not generated by a totally positive element and therefore the function f vanishes on this ideal. □

LEMMA 5.8. *Let r be a nonnegative integer and let $s = [r/2]$. Then*

$$I_D(2^r) = \begin{cases} \dfrac{2^{s-1}\cdot 3}{7} & \text{if } r = 2s, \\[2ex] \dfrac{2^s}{7} & \text{if } r = 2s+1. \end{cases}$$

PROOF. The fundamental unit of k is given by $\varepsilon = \eta/\eta' = 3 - \sqrt{7}$ and we have $\varepsilon < 1$. According to [van der Geer 1988, § V.8, p. 112], if M is a fractional ideal generated by a totally positive element μ satisfying

$$\varepsilon^2 \leq \mu/\mu' \leq 1,$$

then the value of $f(M)$ is also given by

$$f(M) = \frac{1}{\sqrt{D}}\left(\frac{\mu}{1-\varepsilon} + \frac{\mu'\varepsilon}{1-\varepsilon}\right) = \operatorname{tr}\left(\frac{\mu}{(1-\varepsilon)\sqrt{D}}\right).$$

Since 2 is ramified in k, there is only one ideal of k of norm 2^r. If $r = 2s$ the ideal is generated by $\mu = 2^s$. If $r = 2s+1$ the ideal is generated by $\mu = 2^s\cdot\eta$. In either case, $\varepsilon^2 \le \mu/\mu' \le 1$. Therefore,

$$I_D(2^r) = \begin{cases} \operatorname{tr}\left(\dfrac{2^s}{(3\sqrt{7}-7)\sqrt{28}}\right) & \text{if } r = 2s, \\ \operatorname{tr}\left(\dfrac{2^s\cdot\eta}{(3\sqrt{7}-7)\sqrt{28}}\right) & \text{if } r = 2s+1, \end{cases} = \begin{cases} \dfrac{2^{s-1}\cdot 3}{7} & \text{if } r = 2s, \\ \dfrac{2^s}{7} & \text{if } r = 2s+1. \end{cases}$$

We now use Hausmann's formula to compute some intersection numbers on $\widehat{X}$ and $\widehat{X}(\sqrt{7})$. This is entirely straightforward, using the numerical values we have provided for the various number theoretic functions that appear in it and we leave the details to the reader. However, there is one detail that we wish to point out. Hausmann's formula is for intersections of the modular curves $T_M(\widehat{X})$ and $T_N(\widehat{X})$, whereas the table in the following lemma is for intersections of the modular curves $F_M(\widehat{X})$ and $F_N(\widehat{X})$. For $M = 1, 2, 14, 28$, we have $F_M(\widehat{X}) = T_M(\widehat{X})$ and for $M = 4$ we have

$$F_4(\widehat{X}) = T_4(\widehat{X}) - T_1(\widehat{X}).$$

Therefore, Hausmann's formula is perfectly adequate for our purposes. In general, one can always write a curve $F_M(\widehat{X})$ as a linear combination of curves $T_N(\widehat{X})$. □

LEMMA 5.9. *We have the following table of intersection numbers on* $\widehat{X}(\sqrt{7})$.

	$F_1(\widehat{X}(\sqrt{7}))$	$F_2(\widehat{X}(\sqrt{7}))$	$F_4(\widehat{X}(\sqrt{7}))$	$F_{14}(\widehat{X}(\sqrt{7}))$	$F_{28}(\widehat{X}(\sqrt{7}))$
$F_1(\widehat{X}(\sqrt{7}))$	8	24	36	0	84
$F_2(\widehat{X}(\sqrt{7}))$	24	30	24	42	84
$F_4(\widehat{X}(\sqrt{7}))$	36	24	−6	84	42
$F_{14}(\widehat{X}(\sqrt{7}))$	0	42	84	−42	168
$F_{28}(\widehat{X}(\sqrt{7}))$	84	84	42	168	210

The curves $F_N(\widehat{X}(\sqrt{7}))$ are irreducible for $N = 1, 2, 4$ but are reducible for $N = 14$ and $N = 28$. We also need to know the intersections of the individual components of the curves $F_N(\widehat{X}(\sqrt{7}))$ with the curves $F_M(\widehat{X}(\sqrt{7}))$ for $N = 14, 28$ and $M = 1, 2, 4, 14, 28$. Just from knowing the number of components of the curves $F_N(\widehat{X}(\sqrt{7}))$ with $N = 14, 28$ and the above table of intersection numbers, we can obtain some partial information, which we summarize in the following lemma.

COROLLARY 5.10. *Let $Z_N(\widehat{X}(\sqrt{7}))$ denote a component of $F_N(\widehat{X}(\sqrt{7}))$, with $N = 14$ or $N = 28$. Then we have the following intersection numbers*:

	$F_1(\widehat{X}(\sqrt{7}))$	$F_2(\widehat{X}(\sqrt{7}))$	$F_4(\widehat{X}(\sqrt{7}))$	$F_{14}(\widehat{X}(\sqrt{7}))$	$F_{28}(\widehat{X}(\sqrt{7}))$
$Z_{14}(\widehat{X}(\sqrt{7}))$	0	2	4	−2	8
$Z_{28}(\widehat{X}(\sqrt{7}))$	4	4	2	−8	10

The more detailed problem of the intersection of the individual components of $F_{14}(\widehat{X}(\sqrt{7}))$ with themselves, with each other and with those of $F_{28}(\widehat{X}(\sqrt{7}))$ will be taken up in Section 7.

6. The Switching Involution τ

Denote by τ the mapping of $\mathcal{H}^2$ to itself defined by

$$\tau(z_1, z_2) = (z_2, z_1).$$

We will call τ the *switching involution.* One can show that if γ belongs to $\widehat{\Gamma}$ then as mappings of $\mathcal{H}^2$ to itself we have

$$\gamma\tau = \tau\gamma',$$

where the prime indicates that the nontrivial automorphism of k is to be applied to all of the entries of γ. Since the ideal $\sqrt{7}\mathcal{O}_k$ is invariant under the Galois group of k over $\mathbb{Q}$, the group $\widehat{\Gamma}(\sqrt{7})$ is invariant under conjugation by τ.

The proof of the following lemma is straightforward and is left to the reader.

LEMMA 6.1. *Let γ be a matrix with entries in k and totally positive determinant. Write*

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then for $z_1, z_2 \in \mathcal{H}$, the relation

$$z_2 = \gamma \cdot z_1$$

is equivalent to the relation

$$(z_2\ 1)\ B \begin{pmatrix} z_1 \\ 1 \end{pmatrix} = 0,$$

where

$$B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \gamma.$$

It is important for us to know the fixed point set of τ acting on $\widehat{\Gamma}(\sqrt{7})\backslash\mathcal{H}^2$.

LEMMA 6.2. *The fixed point set of τ on the surface $\widehat{\Gamma}(\sqrt{7})\backslash\mathcal{H}^2$ is the union of $F_1(\widehat{\Gamma}(\sqrt{7})\backslash\mathcal{H}^2)$, $F_2(\widehat{\Gamma}(\sqrt{7})\backslash\mathcal{H}^2)$ and $F_4(\widehat{\Gamma}(\sqrt{7})\backslash\mathcal{H}^2)$.*

PROOF. Let $z = (z_1, z_2)$ be a point of $\mathcal{H}^2$. Then z represents a point x of X fixed by τ if and only if there is an element γ of $\Gamma(\sqrt{7})$ such that $\tau z = \gamma z$ or, what is the same,

$$z_2 = \gamma z_1, \quad z_1 = \gamma' z_2.$$

Then we have

$$z_2 = \gamma\gamma' z_2,$$

which implies that $\gamma\gamma'$ has z as a fixed point on $\mathcal{H}^2$. This implies that $\gamma\gamma'$ acts as the identity transformation on $\mathcal{H}^2$. We may assume that the determinant of γ is either 1 or ε. Therefore we can write

$$\gamma = \begin{pmatrix} \xi & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

where the second matrix on the right-hand side, which we will denote δ, has determinant 1 and entries in $\mathcal{O}_k$, and where ξ is the determinant of γ. Note that both matrices on the right-hand side lie in Γ. Since $\xi\xi' = 1$, it follows that

$$\pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \gamma\gamma' = \begin{pmatrix} a & \xi b \\ \xi' c & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

or, what is the same,

$$\begin{pmatrix} a & \xi b \\ \xi' c & d \end{pmatrix} = \pm \begin{pmatrix} d' & -b' \\ -c' & a' \end{pmatrix}.$$

We first remark that the sign on the right-hand side must be $+$. For if it were $-$, we would have $d = -a'$ and therefore modulo $\sqrt{7}$ we would have (since $\varepsilon \equiv 1$ modulo $\sqrt{7}$):

$$1 \equiv \det(\gamma) \equiv ad - bc \equiv -N(a) - bc.$$

This is a contradiction since b, c are congruent to 0 modulo $\sqrt{7}$ and a, a' are congruent to 1 modulo $\sqrt{7}$, hence $N(a)$ is congruent to 1 modulo 7. There are therefore two cases:

(1) the sign is $+$ and $\xi = 1$;
(2) the sign is $+$ and $\xi = \varepsilon$.

In case (1), we have

$$a' = d, \quad b' = -b, \quad c' = -c,$$

and therefore

$$\gamma = \begin{pmatrix} a & b_0\sqrt{7} \\ c_0\sqrt{7} & a' \end{pmatrix}$$

with b_0, c_0 rational integers and a in $\mathcal{O}_k$. The relation

$$z_2 = \gamma z_1 = \frac{az_1 + b}{cz_1 + d} = \frac{az_1 + b_0\sqrt{7}}{c_0\sqrt{7}z_1 + a}$$

then becomes

$$(z_2\ 1)\ B_0 \begin{pmatrix} z_1 \\ 1 \end{pmatrix} = 0,$$

where

$$B_0 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \gamma = \begin{pmatrix} c & d \\ -a & -b \end{pmatrix} = \begin{pmatrix} c_0\sqrt{7} & a' \\ -a & -b_0\sqrt{7} \end{pmatrix}.$$

Thus, the matrix B_0 is skew-hermitian. If b_0, c_0 are even, it is actually an integral skew-hermitian matrix B. If not, we can multiply B_0 by 2 to obtain an integral skew-hermitian matrix B. Since

$$1 = \det(\gamma) = N(a) - 7b_0c_0,$$

there can be no natural number $n > 1$ dividing a, b_0 and c_0. Therefore, the resulting skew-hermitian form will be primitive. If b_0, c_0 are both odd, the determinant of B will be 1. Otherwise, it will be 4. It follows that in case 1, the point z lies in $F_1 \cup F_4$.

In case (2), we have

$$a' = d, \quad b' = -b\varepsilon, \quad c' = -c\varepsilon'.$$

This implies that we can write

$$b = b_0\sqrt{7}(3+\sqrt{7}) = b_0\eta'\sqrt{7},$$
$$c = c_0\sqrt{7}(3-\sqrt{7}) = c_0\eta\sqrt{7},$$

where b_0, c_0 are rational integers and where we recall that $\eta = 3-\sqrt{7}$. Therefore, $z = (z_1, z_2)$ lies in the locus defined by

$$(z_2\ 1)\ B \begin{pmatrix} z_1 \\ 1 \end{pmatrix} = 0,$$

where

$$B = (3+\sqrt{7}) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \gamma = \begin{pmatrix} c_0\sqrt{28} & a'\eta' \\ -a\eta & -b_0\sqrt{28} \end{pmatrix}.$$

The matrix B is integral and skew-hermitian. In fact, B is primitive as well. For if $n > 1$ is a positive integer dividing b_0, c_0, and $a\eta$, then n divides b, c and the identity $\det(\gamma) = \varepsilon$ implies n doesn't divide a. Therefore, since n divides $a\eta$ and η has norm 2, we must have $n = 2$. But then

$$2 = \det(B) = -28b_0c_0 + 2N(a) \equiv 0 \pmod 4,$$

which is a contradiction. It follows that z lies in $F_B \subset F_2$.

We have therefore proved that the fixed point set of τ lies in $F_1 \cup F_2 \cup F_4$. Next we show that the fixed point set contains a component of each of F_1, F_2, F_4. Indeed, the 3 loci

$$z_1 = z_2, \quad z_1 = \varepsilon z_2, \quad z_1 = z_2 + \sqrt{7} \tag{6.3}$$

are easily seen to be invariant under τ modulo $\widehat{\Gamma}(\sqrt{7})$. They are are defined by

$$(z_2\ 1)\ B \begin{pmatrix} z_1 \\ 1 \end{pmatrix} = 0$$

with B given by

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & \eta' \\ \eta & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 2 \\ -2 & \sqrt{7} \end{pmatrix},$$

respectively, and these three matrices are primitive, integral skew-hermitian matrices of determinants 1,2, and 4 respectively. The proof is now completed by the observation that the loci $F_1(\widehat{X}(\sqrt{7}))$, $F_2(\widehat{X}(\sqrt{7}))$, $F_4(\widehat{X}(\sqrt{7}))$ are all irreducible by Lemma 3.3. $\square$

LEMMA 6.4. *The involution τ of $\widehat{\Gamma}(\sqrt{7})\backslash\mathcal{H}^2$ extends to $\widehat{X}(\sqrt{7})$. The fixed point set of τ on the surface $\widehat{X}(\sqrt{7})$ is the union of $F_1(\widehat{X}(\sqrt{7}))$, $F_2(\widehat{X}(\sqrt{7}))$ and $F_4(\widehat{X}(\sqrt{7}))$.*

PROOF. The first assertion is obvious from the definitions. As for the second, it follows from Lemma 6.2, from Lemma 3.6 and from the fact that all of the cusps of $\widehat{X}(\sqrt{7})$ are rational. $\square$

7. Intersections on the Nonsingular Model $\widehat{Z}(\sqrt{7})$ of $\widehat{X}(\sqrt{7})$

Hirzebruch [1971] showed how to resolve the singularities of Hilbert modular surfaces. His construction inspired the more general work of Mumford and others [Ash et al. 1975, § I.5, pp. 39–53] on compactification and desingularization of quotients of Hermitian symmetric spaces by arithmetic groups. We will discuss some aspects of this more general work in [Adler $\geq$ 1998].

The only singularities of $\widehat{X}(\sqrt{7})$ are at the cusps. Applying Hirzebruch's construction to the surface $\widehat{X}(\sqrt{7})$, we obtain a surface which we will denote $\widehat{Z}(\sqrt{7})$. The preimage in $\widehat{Z}(\sqrt{7})$ of each cusp is a pentagonal cycle of 5 rational curves with self intersection numbers -2, -2, -3, -3, -2 respectively. This is described in [Hirzebruch 1977]. It can be justified from the general theory of such resolutions by means of the table in [van der Geer 1988, p. 41], where we take $\alpha = 3$ in the second row, second column of the table. We illustrate the resolution cycle in Figure 1.

If x is a cusp of $\widehat{X}(\sqrt{7})$, we will denote by S_i^x, for $0 \leq i \leq 4$, the 5 curves of the resolution cycle. We number them with subscripts modulo 5 such that

(1) The curve S_i^x has self-intersection number -2 for $i = 0, 1, 4$ and self-intersection number -3 for $i = 2, 3$.

(2) The curves S_i^x and S_j^x meet transversely in one point if $i - j$ is congruent to ± 1 modulo 5 and otherwise do not meet.

Our main goal in this section is to study the intersection numbers of some of the curves $F_N(\widehat{Z}(\sqrt{7}))$ on $\widehat{Z}(\sqrt{7})$. The key to computing these numbers is the

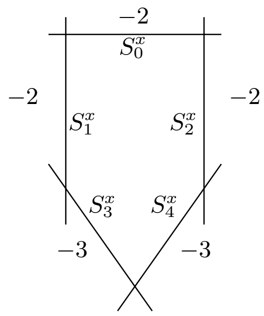


Figure 1. Resolution cycle of cusp x in $\widehat{Z}(\sqrt{7})$.

computation of the corresponding numbers on $\widehat{X}(\sqrt{7})$, which was carried out in Section 5, combined with Mumford's definition [Mu1] of intersection numbers on complete normal singular surfaces. Therefore, although we merely alluded to that notion at the beginning of Section 5, here we need to consider it explicitly and we will now do so.

Let S be a complete surface with isolated normal singularities. Denote by S' a surface obtained by resolving the singularities of S and denote by $f : S' \to S$ the natural mapping. If C is a curve on S, we will denote by C' the preimage of C on S'. For each point p of S such that $f^{-1}(p)$ is a curve, denote the components of $f^{-1}(p)$ by $K_1^p, \ldots, K_r^p$, where $r = r(p)$ depends on p. Denote by M_p the matrix $K_i^p \cdot K_j^p$ of intersection numbers of the components of $f^{-1}(p)$. Since $f^{-1}(p)$ can be blown down to a point, namely p, the matrix M_p is negative definite and, in particular, invertible as a matrix with rational entries. If C is a curve on S, denote by $\overline{C} = \overline{C}(f)$ the cycle with rational coefficients in S' given by

$$\overline{C} = C' + \sum_p \sum_{i=1}^{r(p)} a_{ip} K_i^p,$$

where the outer summation runs over all points p of S such that $f^{-1}(p)$ is a curve and where, for each such p, the a_{ip} are the unique rational numbers such that

$$\overline{C} \cdot K_i^p = 0$$

for $1 \leq i \leq r(p)$. One then defines the intersection number of two curves C_1, C_2 on S to be the intersection number on S' of the curves $\overline{C}_1$, $\overline{C}_2$:

$$(C_1 \cdot C_2)_S = (\overline{C}_1 \cdot \overline{C}_2)_{S'}.$$

One can show that the intersection number so defined is independent of the choice of the resolution $f : S' \to S$. We also note that since the preimage of a curve on S differs from the proper transform of the curve only by an integral linear combination of curves of the form K_i^p, one could have taken C' in the above definition to be be the proper transform of C or, indeed, any curve in S' mapping onto C.

If $S' \to S$ is a morphism of surfaces of Hilbert modular type, with S' nonsingular and $C = F_N(S)$, we will also write $F_N(S' \to S)$ to denote the curve $\overline{C}$ on S'.

Lemma 7.1. *We have*

$$F_1(\widehat{Z}(\sqrt{7}) \to \widehat{X}(\sqrt{7})) = F_1(\widehat{Z}(\sqrt{7})) + \frac{1}{2}\sum_x (3S_0^x + 2S_1^x + S_2^x + S_3^x + 2S_4^x)$$

$$F_2(\widehat{Z}(\sqrt{7}) \to \widehat{X}(\sqrt{7})) = F_2(\widehat{Z}(\sqrt{7})) + \sum_x (S_0^x + S_1^x + S_2^x + S_3^x + S_4^x)$$

$$F_4(\widehat{Z}(\sqrt{7}) \to \widehat{X}(\sqrt{7})) = F_4(\widehat{Z}(\sqrt{7})) + \frac{1}{2}\sum_x (3S_0^x + 2S_1^x + S_2^x + S_3^x + 2S_4^x).$$

Proof. To prove this, one verifies that the intersection of a component S_i^x with the expression on the right side of each of these equations is 0. Since we know the intersection numbers $S_i^x \cdot S_j^y$, one merely needs to know the intersection numbers $F_N(\widehat{Z}(\sqrt{7}))\cdot S_i^x$. They are given by

$$F_N(Z(\sqrt{7}))\cdot S_i^x = \begin{cases} 1 & \text{if } (N,i) = (1,0),\ (2,2),\ (2,3),\ (4,0), \\ 0 & \text{otherwise.} \end{cases}$$

That these are the correct intersection numbers between the $F_N(Z(\sqrt{7}))$ and the S_i^x will be shown in Lemma 7.7. □

In order to prepare for the proof of Lemma 7.7, we need to look more closely at Hirzebruch's desingularization of $\widehat{X}(\sqrt{7})$.

A sequence indexed by the set of all integers will be called a $\mathbb{Z}$-*sequence*. If $\boldsymbol{b} = (b_n)_{n\in\mathbb{Z}}$ is a $\mathbb{Z}$-sequence of positive integers, one can associate to $\boldsymbol{b}$ a complex manifold which we will denote $\mathcal{M}_0(\boldsymbol{b})$. The explicit construction is described in [Hirzebruch 1973; 1971; van der Geer 1988]. For certain purposes, one assumes that all of the integers b_n are ≥ 2 with at least one of them ≥ 3, and that is the practice of the authors we have cited. One constructs $\mathcal{M}_0(\boldsymbol{b})$ by forming the disjoint union of a $\mathbb{Z}$-sequence of copies of $\mathbb{C}^2$ and identifying the point (u_n, v_n) in the n-th copy $\mathbb{C}^2_n$ with the point $(u_n^{b_n} v^n, u_n^{-1})$ in the $(n+1)$-th copy $\mathbb{C}^2_{n+1}$. Each copy of $\mathbb{C}^2$ injects into $\mathcal{M}_0(\boldsymbol{b})$ and the injection defines a coordinate chart $(\mathbb{C}^2_n, (u_n, v_n))$. The closure in $\mathcal{M}_0(\boldsymbol{b})$ of the axis $v_n = 0$ of that chart will be a projective line which we will denote S_n and which has self-intersection number $-b_n$ (cf. [van der Geer 1988, p. 33]).

In the following lemma, it is convenient to denote by $\mathcal{I}(\mathbb{Z})$ the isometry group of the the set of integers with their usual metric inherited from the real numbers. As is well-known, every such isometry is of the form $n \mapsto tn + r$, where $r \in \mathbb{Z}$ and $t = \pm 1$. In order to conserve subscripts, in the proof of the lemma we will write $(u, v)_n$ for all $n \in \mathbb{Z}$ to denote a point whose coordinates are (u, v) in the coordinate chart σ_n.

LEMMA 7.2. *Denote by ι the automorphism of $\mathbb{C}^2$ given by $(u,v) \mapsto (v,u)$. Then the isometry group $\mathcal{I}(\mathbb{Z})$ of $\mathbb{Z}$ acts holomorphically on the disjoint union $\mathbb{C} \times \mathbb{Z}$ of copies of $\mathbb{C}^2$ indexed by $\mathbb{Z}$ by the rule*

$$\alpha_{(t,r)} \cdot (w,n) = \left(\iota^{(1-t)/2} w,\, t(n - \tfrac{1}{2}) + \tfrac{\pm}{r}\right),$$

with $t = \pm 1$, where $\alpha_{(t,r)}$ is the automorphism of $\mathbb{Z}$ given by $n \mapsto tn + r$. Let $\boldsymbol{b} : \mathbb{Z} \to \mathbb{Z}_{\geq 1}$ be a $\mathbb{Z}$-sequence of positive integers. Let $\mathcal{G}$ be the subgroup of $\mathcal{I}(\mathbb{Z})$ consisting of all automorphisms g of $\mathbb{Z}$ which preserve $\boldsymbol{b}$ in the sense that $\boldsymbol{b} \circ g = \boldsymbol{b}$. Then the action of $\mathcal{G}$ on the complex manifold $\mathbb{C}^2 \times \mathbb{Z}$ descends to an action of $\mathcal{G}$ on the complex manifold $\mathcal{M}_0(\boldsymbol{b})$.

PROOF. That we have indeed defined an action of $\mathcal{I}(\mathbb{Z})$ on $\mathbb{C}^2 \times \mathbb{Z}$ as a group of holomorphic automorphisms is straightforward and left to the reader. As for the assertion that $\mathcal{G}$ acts on $\mathcal{M}_0(\boldsymbol{b})$, it suffices to show that for $g \in \mathcal{G}$, the action of g on $\mathbb{C}^2 \times \mathbb{Z}$ is compatible with the identifications which give rise to the complex manifold $\mathcal{M}_0(\boldsymbol{b})$. To see this, let $p = (u,v)_n$ be a point of $\mathbb{C}^2_n$. This point is identified with the point $q = (u^{b_n} v, 1/u)_{n+1}$ of $\mathbb{C}^2_{n+1}$. Let $\alpha_{(t,r)}$ be an element of $\mathcal{G}$. Let $p' = \alpha_{(t,r)}(p)$ and $q' = \alpha_{(t,r)}(q)$. If $t = 1$, we have

$$p' = (u,v)_{n+r}, \quad q' = (u^{b_n} v,\, 1/u)_{n+r+1},$$

and p' is identified with

$$(u^{b_{n+r}} v, 1/u)_{n+r+1} = (u^{b_n} v, 1/u)_{n+r+1} = q'$$

since $b_{n+r} = b_n$. This proves that the translations in $\mathcal{G}$ descend to $\mathcal{M}_0(\boldsymbol{b})$. Therefore, it suffices to show that if $\alpha_{(-1,0)}$ belongs to $\mathcal{G}$ then it descends to $\mathcal{M}_0(\boldsymbol{b})$. Hence we may assume that $b_{-n} = b_n$ for all n. Then as above, we have

$$p' = \alpha_{(-1,0)}(p) = (v,u)_{1-n}, \quad q' = \alpha_{(-1,0)}(q) = (1/u,\, u^{b_n} v)_{-n},$$

and q' is identified with

$$((1/u)^{b_{-n}} (u^{b_n} v), 1/(1/u))_{1-n} = (v,u)_{1-n} = p'$$

since $b_{-n} = b_n$. This proves the lemma. □

In most of the cases of interest in this theory, the sequence is periodic and in many cases it is also symmetric under $n \mapsto -n$. If $\boldsymbol{b}$ is periodic with period r, where r is not necessarily the smallest period of $\boldsymbol{b}$, then $r\mathbb{Z}$ fixes $\boldsymbol{b}$ and therefore acts on $\mathcal{M}_0(\boldsymbol{b})$. The quotient of $\mathcal{M}_0(\boldsymbol{b})$ for the action of $r\mathbb{Z}$ will be denoted $\mathcal{M}_r(\boldsymbol{b})$, and $\mathcal{M}_0(\boldsymbol{b})$ may be regarded as the special case $r = 0$. The images in $\mathcal{M}_r(\boldsymbol{b})$ of the curves S_n will form a closed cycle of rational curves. In case $\boldsymbol{b}$ is not only periodic but also invariant under $n \mapsto -n$, the latter induces an automorphism of $\mathcal{M}_r(\boldsymbol{b})$ which we will denote $\tau_{r,\boldsymbol{b}}$. This choice of notation derives from the relation between the automorphism $\tau_{r,\boldsymbol{b}}$ and the switching map τ, which will be elucidated below.

COROLLARY 7.3. *Assume that $\boldsymbol{b}$ is periodic with period r and symmetric under $n \mapsto -n$. Then the involution $\tau_{r,\boldsymbol{b}}$ interchanges the the projective lines $S_n(\boldsymbol{b})$ and $S_{-n}(\boldsymbol{b})$ for all n.*

PROOF. The component $S_n(\boldsymbol{b})$ is the closure of the axis $v_n = 0$ of the n-th coordinate system. It is mapped to the closure of the axis $u_{1-n} = 0$ in the $(1-n)$-th coordinate system. But that is the same as the closure of the axis $v_{-n} = 0$ of the n-th coordinate system, which is $S_{-n}(\boldsymbol{b})$, and we are done. □

The manifolds $\mathcal{M}_r(\boldsymbol{b})$ are the key to Hirzebruch's resolution of the cusp singularities of Hilbert modular surfaces. We refer the reader to the works cited above for details. We will merely summarize the facts that are pertinent to our work here.

Let Γ be a group of Hilbert modular type acting on $\mathcal{H}^2$ and let X_Γ denote the SBB compactification of $\Gamma\backslash\mathcal{H}^2$. Modulo ± 1, the subgroup fixing the cusp ∞ will be the semidirect product of a subgroup V of finite index in the group of totally positive units of $\mathcal{O}_k$ with a lattice M in k. It is convenient to denote the subgroup fixing ∞ by (V, M). The norm form of the lattice gives rise to a rational binary quadratic form, a root of which can be expanded in a continued fraction of Hirzebruch's type,

$$c_0 - \cfrac{1}{c_1 - \cfrac{1}{c_2 - \cfrac{1}{c_3 - \dots}}}.$$

Since the root is a quadratic irrationality, the coefficients c_n are eventually periodic in n. If the smallest period is $(b_1 \dots b_r)$ then one forms the $\mathbb{Z}$-sequence $\boldsymbol{b} = (b_n)_{n\in\mathbb{Z}}$ by extending b_n to a periodic function of n with period r on all of $\mathbb{Z}$. The resolution of the singularity at the cusp ∞ in the SBB compactification of $\Gamma\backslash\mathcal{H}^2$ is then obtained [Hirzebruch 1971, Theorem of § 3] by replacing a neighborhood of the singularity by a neighborhood of the union of the curves curves S_n in $\mathcal{M}_{rs}(\boldsymbol{b})$, where s is the index of V in the full unit group of $\mathcal{O}_k$ modulo ± 1. (Some care has to be taken in the case where $rs \leq 2$; see [Hirzebruch 1971] for details.) The components of the resolution cycle at a cusp x will be denoted S_0^x, S_1^x, ..., if they are known from the context, or by $S_0^x(\delta)$, $S_1^x(\delta)$, etc., where δ is data such as $\mathcal{O}_k$, $\sqrt{7}\mathcal{O}_k$ or $\boldsymbol{b}$ describing the group Γ and the resolution in more or less detail.

One can be quite explicit about this construction. Indeed, if one embeds the lattice M into $\mathbb{R}^2$ via the two embeddings of k into $\mathbb{R}$, the image of the set of totally positive elements of M will be denoted M_+. The convex hull of M_+ will be an infinite polygon whose vertices form a $\mathbb{Z}$-sequence A_n in k. If we regard k as a subfield of $\mathbb{R}$ by fixing one of the embeddings in advance, then we can determine the $\mathbb{Z}$-sequence A_n up to a translation of the index n by requiring that

A_n be monotonically decreasing in n. The $\mathbb{Z}$-sequence A_n then has the following interesting properties for all n:

(1) A_{n-1} and A_n form a basis for M;
(2) $A_{n-1} + A_{n+1} = b_n A_n$.

Denoting by B_n, C_n the dual basis to A_{n-1}, A_n with respect to the trace form $(x, y) \mapsto \mathrm{tr}(xy)$ of k, we can then define a map

$$\phi_n : \mathcal{H}^2 \to \mathbb{C}_n^2 \subseteq \mathcal{M}_0(\boldsymbol{b})$$

by the rule

$$\phi_n(z_1, z_2) = \bigl(e(B_n z_1 + B_n' z_2), e(C_n z_1 + C_n' z_2)\bigr) = (u_n, v_n),$$

where $e(t) = e^{2\pi i t}$ for all t. One can then verify that for z, w in $\mathcal{H}^2$, we have $\phi_n(z) = \phi_n(w)$ if and only if for some $\mu \in M$ we have

$$w = \begin{pmatrix} 1 & \mu \\ 0 & 1 \end{pmatrix} z = (z_1 + \mu, z_2 + \mu'),$$

where $z = (z_1, z_2)$. Furthermore, $\eta = \varepsilon^s$ is a generator of V and we have

$$\phi_{n+rs} \circ \begin{pmatrix} \eta & 0 \\ 0 & 1 \end{pmatrix} \cdot z = \phi_{n+rs}(\eta z_1, \eta' z_2) = \phi_n(z)$$

for all n and all $z = (z_1, z_2)$. Therefore, ϕ_n induces a map

$$(V, M)\backslash\mathcal{H}^2 \to \mathbb{C}_n^2 \subseteq \mathcal{M}_{rs}(\boldsymbol{b})$$

which is injective in a neighborhood of ∞ and maps that neighborhood isomorphically onto a neighborhood in $\mathcal{M}_{rs}(\boldsymbol{b})$ of the form $W - \cup S_n$, where W is a neighborhood of $\cup S_n$ in $\mathcal{M}_{rs}(\boldsymbol{b})$. The inverse mapping extends to a mapping from W to a neighborhood of ∞ in the SBB compactification and that mapping resolves the singularity at ∞. The preimage of ∞ will be called the *resolution cycle* at ∞. In the special case where $rs \leq 2$, some care must be taken, as described in [Hirzebruch 1971, § 3], but we will not worry about such details here. Using elements of $GL_2^+(k)$, one can move any cusp to ∞ and back. Therefore, one can give a similar description of the resolution of the cusps of the SBB compactification of $\Gamma\backslash\mathcal{H}^2$ for any group Γ of Hilbert modular type. In the particular case we are considering, where $k = \mathbb{Q}(\sqrt{7})$, the field k has class number 1, and $\Gamma = \widehat{\Gamma}(\sqrt{7})$, the cusps are all equivalent under the action of $\widehat{\Gamma}$ and the description of the resolution is simpler. In particular, all of the resolution cycles for X will look the same. The surface we have denoted by $\widehat{Z}(\sqrt{7})$ has been obtained by this process.

Thanks to the system of bases A_{n-1}, A_n and the associated maps ϕ_n, we have a very clear picture of what $\widehat{Z}(\sqrt{7})$ looks like in a neighborhood of the resolution cycles of the cusps. In this case, we can compute the system of bases A_{n-1}, A_n quite explicitly, thanks to the table in [van der Geer 1988, p. 41], which gives for the the cycle associated to the lattice $\sqrt{28}\mathcal{O}_k$ (the same as the one associated with $\sqrt{7}\mathcal{O}_k$, since the two lattices are related by multiplication by a rational number) as $(3, 3, 2, 2, 2)$. However, we find it more convenient to describe the cycle as $(2, 2, 3, 3, 2)$.

LEMMA 7.4. *The* $\mathbb{Z}$*-sequence* $A_n = A_n(\sqrt{7}\mathcal{O}_k)$ *is given by*

$$A_n(\sqrt{p}\mathcal{O}_k) = \varepsilon^m(7 - j\sqrt{p})$$

if $n = j+5m$, *with* $|j| \le 2$, *where* $m \in \mathbb{Z}$ *and where the corresponding* $\mathbb{Z}$*-sequence* $\{b_n\}$ *is given by*

$$b_n = \begin{cases} 3 & \text{if } n \equiv \pm 2 \pmod 5, \\ 2 & \text{otherwise.} \end{cases} \tag{7.5}$$

PROOF. It is straightforward to verify that with $M = \sqrt{7}\mathcal{O}_k$ and $A_n = A_n(M)$ and b_n as given above, we have

$$A_{n-1} + A_{n+1} = b_n A_n$$

for all n in $\mathbb{Z}$. Furthermore, the $\mathbb{Z}$-sequence b_n is cyclic with period $(3, 3, 2, 2, 2)$ if $M = \sqrt{7}\mathcal{O}_k$. Furthermore, we have $A_{-n} = A'_n$ and $b_{-n} = b_n$ for all $n \in \mathbb{Z}$. Furthermore, A_0, A_1 is a basis for M, namely $7, 7 - \sqrt{7}$ if $M = \sqrt{7}\mathcal{O}_k$. Therefore the A_n are precisely the vertices of the convex hull of the set M_+ of totally positive elements of M [van der Geer 1988, pp. 31–33]. □

LEMMA 7.6. *Let* x *be a cusp of* $\widehat{X}(\sqrt{7})$. *The only fixed points of* τ *on the components* S^x_i *of the resolution cycles of the cusps are its fixed points on the components* S^x_0 *and the point where the two components* S^x_2 *and* S^x_3 *meet. More precisely, since* τ *induces an involution of the nonsingular rational curve* S^x_0, *it will have two fixed points on* S^x_0 *and therefore* 3 *fixed points in all among the cusps.*

PROOF. Since the period of $(2, 2, 3, 3, 2)$ is 5 and since the order of M is the maximal order of k, the value of s is 1 in this case. Therefore, a neighborhood of each cusp x of $\widehat{X}(\sqrt{7})$ is replaced by a neighborhood of the cycle of curves $S^x_i(M)$, $0 \le i \le 4$, in the complex manifold $\mathcal{M}_5\overline{(2, 2, 3, 3, 2)}$. Furthermore, it is clear from the construction of $\mathcal{M}_5\overline{(2, 2, 3, 3, 2)}$, and from the fact that G acts transitively on the set of cusps and the fact that all of the cusps of $\widehat{\Gamma}(\sqrt{7})$ are rational, that the involution τ of $\widehat{\Gamma}(\sqrt{7})\backslash\mathcal{H}^2$ extends to an automorphism, also denoted τ, of $\widehat{Z}(\sqrt{7})$ inducing the automorphism $\tau_{5,\boldsymbol{b}}$, where $\boldsymbol{b} = \overline{(2, 2, 3, 3, 2)}$ on the copy of $\mathcal{M}_5(\boldsymbol{b})$ at each cusp. It therefore follows from Corollary 7.3 that τ interchanges the curves S^x_i and S^x_{-i}. From this, the lemma follows at once. □

We know from Lemma 6.4 that the involution τ acting on $\widehat{X}(\sqrt{7})$ has for its fixed point set the union of the three curves $F_N(\widehat{X}(\sqrt{7}))$, with $N = 1, 2, 4$. Therefore, the fixed point set of τ acting on $\widehat{Z}(\sqrt{7})$ will contain the union of the three curves $F_N(\widehat{Z}(\sqrt{7}))$. The following lemma sharpens this observation and also determines the intersection number of each $F_N(\widehat{X}(\sqrt{7}))$, for $N = 1, 2, 4$ with the components S_i^x.

LEMMA 7.7. *Let x be a cusp of $\widehat{X}(\sqrt{7})$. Each of the three fixed points of τ among the five components S_i^x, $0 \le i \le 4$, of the resolution cycle of x lies in one and only one of the curves $F_N(\widehat{Z}(\sqrt{7}))$, $N = 1, 2, 4$. The point of intersection of S_2^x and S_3^x lies on $F_2(\widehat{Z}(\sqrt{7}))$. The other two, which are found on S_0^x, lie on $F_1(\widehat{Z}(\sqrt{7}))$ and $F_4(\widehat{Z}(\sqrt{7}))$ respectively. The curves $F_N(\widehat{Z}(\sqrt{7}))$, with $N = 1, 2, 4$, do not have any other points of intersection with the components S_i^x, $0 \le i \le 4$. Furthermore, the intersections of $F_1(\widehat{Z}(\sqrt{7}))$ and $F_4(\widehat{Z}(\sqrt{7}))$ with S_0^x and of $F_2(\widehat{Z}(\sqrt{7}))$ with S_2^x and S_3^x are transverse.*

PROOF. We prove this by looking at the equations of $F_N(\widehat{X}(\sqrt{7}))$, for $N = 1, 2, 4$, in local coordinate charts σ_i containing the fixed points of τ acting on the components S_i^x. Since G acts transitively on the cusps, we can assume that x is the cusp $\infty = (\infty, \infty)$. The coordinate chart σ_1 does not contain the entire component S_0^x, but the point it does not contain lies on the component S_4^x therefore can't be a fixed point. Therefore, we can use the coordinate chart σ_1 to study the fixed points lying on S_0^x. Since the dual basis B_1, C_1 to A_0, A_1 with respect to the pairing $xy \mapsto \operatorname{tr}(xy)$ is given by $B_1 = (1 + \sqrt{7})/2$, $C_1 = -\sqrt{7}/2$, the coordinates (z_1, z_2) of $\mathcal{H}^2$ are related to the coordinates (u_1, v_1) of σ_1 by

$$(u_1, v_1) = \left(e\Big(\frac{1+\sqrt{7}}{2} z_1 + \frac{1-\sqrt{7}}{2} z_2\Big),\ e\Big(-\frac{z_1 + z_2}{2}\Big)\right).$$

Since the curves $F_N(\widehat{Z}(\sqrt{7}))$, for $N = 1, 2, 4$, have only one component each, we obtain them as the images of the curves of the curves described in (6.3), namely:

$$z_1 = z_2, \quad z_1 = \varepsilon z_2, \quad z_1 = z_2 + \sqrt{7}.$$

These three curves are respectively mapped to the following curves in σ_1:

$$v_1 = 1, \quad u_1^7 v_1^4 = 1, \quad v_1 = -1.$$

For example, if we put $z_1 = z_2$, we obtain $(u_1, v_1) = (e(z_2), 1)$; the other cases are handled similarly. From this it is clear that the curves $F_1(\widehat{Z}(\sqrt{7}))$ and $F_4(\widehat{Z}(\sqrt{7}))$ meet the line S_0^x (i.e., $u_1 = 0$) transversely in the two distinct points $(0, 1)$ and $(0, -1)$, respectively, and the curve $F_2(\widehat{Z}(\sqrt{7}))$ does not meet S_0^x. To complete the proof, we work in the coordinate system σ_3, whose intersections with the curves S_2^x, S_3^x are the coordinate axes of the coordinate system σ_3. The points of S_2^x, S_3^x that do not lie on σ_3 lie on the curves S_1^x and S_4^x respectively and therefore are not fixed points of τ. Therefore it suffices to work in the coordinate system σ_3. By Lemma 7.4, we have

$$A_2 = 7 - 2\sqrt{7}, \quad A_3 = 14 - 5\sqrt{7},$$

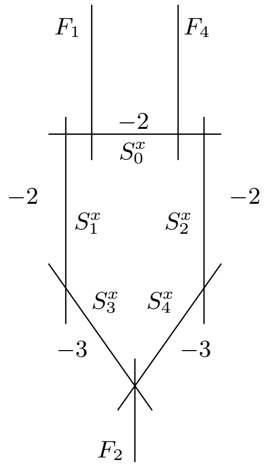


Figure 2. How the curves $F_N(\widehat{Z}(\sqrt{7}))$, for $N = 1, 2, 4$, meet the resolution cycle.

and the dual basis B_3, C_3 to this basis for $\sqrt{7}\mathcal{O}_k$ is given by

$$B_3 = \frac{5+2\sqrt{7}}{14}, \quad C_3 = -\frac{2+\sqrt{7}}{14}.$$

Therefore the coordinates (z_1, z_2) of $\mathcal{H}^2$ are related to the coordinates (u_3, v_3) of σ_3 by

$$(u_3, v_3) = \left(e\left(z_1 \frac{5+2\sqrt{7}}{14} + z_2 \frac{5-2\sqrt{7}}{14} \right),\ e\left(-z_1 \frac{2+\sqrt{7}}{14} - z_2 \frac{2-\sqrt{7}}{14} \right) \right).$$

The curves (6.3) are respectively mapped to the following curves in σ_3:

$$u_3^2 v_3^5 = 1, \quad u_3 = v_3, \quad u_3^2 v_3^5 = -1.$$

Thus, $F_1(\widehat{Z}(\sqrt{7}))$ and $F_4(\widehat{Z}(\sqrt{7}))$ do not meet the axes of σ_3 and $F_2(\widehat{Z}(\sqrt{7}))$ passes through the origin of σ_3 (the point of intersection of S_2^x and S_3^x), transversely to both axes. This completes the proof of the lemma. $\square$

The configuration of curves in the resolution cycle of a cusp x and their intersections with the curves $F_N(\widehat{\Gamma}(\sqrt{7}))$ for $N = 1, 2, 4$ are shown in Figure 2.

Having computed the divisors with rational coefficients $F_N(\widehat{Z}(\sqrt{7}) \to \widehat{X}(\sqrt{7}))$ for $N = 1, 2, 4$, we can now compute the intersection numbers of the modular curves on the surface $\widehat{Z}(\sqrt{7})$. Before doing so, we note that according to Lemma 3.6, the curves $F_{14}(\widehat{X}(\sqrt{7}))$ and $F_{28}(\widehat{X}(\sqrt{7}))$ do not pass through the cusps of $\widehat{X}(\sqrt{7})$. This is also consistent with the fact that by Lemma 2.5 the components of $F_N(\widehat{X}(\sqrt{7}))$ for $N = 14, 28$ are quotients of the upper half plane by arithmetic groups arising from quaternion division algebras over $\mathbb{Q}$ and are therefore compact subvarieties of $\widehat{\Gamma}(\sqrt{7})\backslash\mathcal{H}^2$. Therefore we have

$$F_N(\widehat{Z}(\sqrt{7}) \to \widehat{X}(\sqrt{7})) = F_N(\widehat{Z}(\sqrt{7}))$$

for $N = 14, 28$. That implies that the intersection numbers of $F_N(\widehat{Z}(\sqrt{7}))$ on $\widehat{Z}(\sqrt{7})$ will be the same as the corresponding intersection numbers on $\widehat{X}(\sqrt{7})$. That allows us to fill in the last two rows and columns of the table in the following lemma simply by copying them from the corresponding entries of the table in Lemma 5.9.

LEMMA 7.8. *We have the following table of intersection numbers on $\widehat{Z}(\sqrt{7})$.*

	$F_1(\widehat{Z}(\sqrt{7}))$	$F_2(\widehat{Z}(\sqrt{7}))$	$F_4(\widehat{Z}(\sqrt{7}))$	$F_{14}(\widehat{Z}(\sqrt{7}))$	$F_{28}(\widehat{Z}(\sqrt{7}))$
$F_1(\widehat{Z}(\sqrt{7}))$	28	0	0	0	84
$F_2(\widehat{Z}(\sqrt{7}))$	0	−18	0	42	84
$F_4(\widehat{Z}(\sqrt{7}))$	0	0	−42	84	42
$F_{14}(\widehat{Z}(\sqrt{7}))$	0	42	84	−42	168
$F_{28}(\widehat{Z}(\sqrt{7}))$	84	84	42	168	210

PROOF. According to Mumford's theory, we have

(7.9)
$$F_M(\widehat{X}(\sqrt{7}))\cdot F_N(\widehat{X}(\sqrt{7})) = F_M(\widehat{Z}(\sqrt{7}) \to \widehat{X}(\sqrt{7}))\cdot F_N(\widehat{Z}(\sqrt{7}) \to \widehat{X}(\sqrt{7})).$$

The values of the left side are given by the table in Lemma 5.9 for $M, N = 1, 2, 4, 14, 28$. The factors of the intersection product on the right side of (7.9) can be written in the form

$$F_M(\widehat{Z}(\sqrt{7})) + \sum_x \sum_{i=0}^{4} a_{ixM} S_i^x,$$
$$F_N(\widehat{Z}(\sqrt{7})) + \sum_x \sum_{i=0}^{4} b_{ixN} S_i^x,$$

where the coefficients a_{ixM}, a_{ixN} are known to us from Lemma 7.1. Furthermore, since

$$F_M(\widehat{Z}(\sqrt{7}) \to \widehat{X}(\sqrt{7}))\cdot S_i^x = 0,$$

the intersection product on the right side of (7.9) is equal to

$$\begin{aligned} &F_M(\widehat{Z}(\sqrt{7}) \to \widehat{X}(\sqrt{7}))\cdot F_N(\widehat{Z}(\sqrt{7}) \to \widehat{X}(\sqrt{7})) \\ &\quad = F_M(\widehat{Z}(\sqrt{7}) \to \widehat{X}(\sqrt{7}))\cdot (F_N(\widehat{Z}(\sqrt{7}) + \sum_x \sum_{i=0}^{4} a_{ixN} S_i^x) \\ &\quad = F_M(\widehat{Z}(\sqrt{7}) \to \widehat{X}(\sqrt{7}))\cdot F_N(\widehat{Z}(\sqrt{7})) \\ &\quad = F_M(\widehat{Z}(\sqrt{7}))\cdot F_N(\widehat{Z}(\sqrt{7})) + \sum_x \sum_{i=0}^{4} a_{ixM} S_i^x \cdot F_N(\widehat{Z}(\sqrt{7})). \end{aligned}$$

Knowing the intersection numbers of the modular curve $F_N(\widehat{Z}(\sqrt{7}))$ with the cuspidal components S_i^x, we can use (7.9) to solve for $F_M(\widehat{Z}(\sqrt{7}))\cdot F_N(\widehat{Z}(\sqrt{7}))$ in terms of quantities we know how to compute. In fact, we have

$$(7.10)\quad F_M(\widehat{Z}(\sqrt{7}))\cdot F_N(\widehat{Z}(\sqrt{7})) \\ = F_M(\widehat{X}(\sqrt{7}))\cdot F_N(\widehat{X}(\sqrt{7})) - \sum_x \sum_{i=0}^{4} a_{ixM} S_i^x \cdot F_N(\widehat{Z}(\sqrt{7})).$$

We will leave the detailed verification of the table to the reader, but we will illustrate the computation of one of its entries. According to Lemma 5.9, the self-intersection number of $F_1(\widehat{X}(\sqrt{7}))$ is 8. Using Lemma 7.1 and equation (7.10), we therefore have

$$\begin{aligned} F_1(\widehat{Z}(\sqrt{7}))^2 &= 8 - \frac{1}{2}\sum_x F_1(\widehat{Z}(\sqrt{7}))\cdot(3S_0^x + 2S_1^x + S_2^x + S_3^x + 2S_4^x) \\ &= 8 - \frac{1}{2}\cdot 24\cdot 3 = -28. \end{aligned}$$

□

Like their counterparts on $\widehat{X}(\sqrt{7})$, the curves $F_N(\widehat{Z}(\sqrt{7}))$ are irreducible for $N = 1, 2, 4$ but reducible for $N = 14$ and $N = 28$. As noted above, the two curves for $N = 14, 28$ do not pass through the cusps and there intersection properties are therefore unaffected by the resolution of the cuspidal singularities. The same applies to the irreducible components of these curves. From Corollary 5.10, we therefore immediately have the following result.

COROLLARY 7.11. *Let $Z_N(\widehat{Z}(\sqrt{7}))$ denote a component of $F_N(\widehat{Z}(\sqrt{7}))$, with $N = 14$ or $N = 28$. Then we have the following intersection numbers*:

	$F_1(\widehat{Z}(\sqrt{7}))$	$F_2(\widehat{Z}(\sqrt{7}))$	$F_4(\widehat{Z}(\sqrt{7}))$	$F_{14}(\widehat{Z}(\sqrt{7}))$	$F_{28}(\widehat{Z}(\sqrt{7}))$
$Z_{14}(\widehat{Z}(\sqrt{7}))$	0	2	4	-2	8
$Z_{28}(\widehat{Z}(\sqrt{7}))$	4	4	2	8	10

In order to compute the self-intersection numbers of $Z_N(\widehat{Z}(\sqrt{7}))$ for $N = 14, 28$, we need to recall the adjunction formula and its interpretation for Hilbert modular surfaces.

LEMMA 7.12 (ADJUNCTION FORMULA). *Let W be a complete nonsingular surface and let V be a curve on W. Then we have*

$$2 - 2p_a(V) = c_1(W)\cdot V - V\cdot V,$$

where $p_a(V)$ is the arithmetic genus of V and $c_1(W)$ is the first Chern class of W.

PROOF. See [van der Geer 1988, p. 162]. Here $p_a(V)$ is just the genus of V if V is nonsingular. □

Using the adjunction formula, one can compute self-intersection numbers provided one can compute the other terms in the formula. The next two tools provide formulas for computing $c_1(\widehat{Z}(\sqrt{7}))\cdot F_N(\widehat{Z}(\sqrt{7}))$ and $p_a(F_N(\widehat{Z}(\sqrt{7})))$ respectively.

LEMMA 7.13. *Let B be a primitive integral skew hermitian matrix over k. Then the intersection number of $F_B(\widehat{Z}(\sqrt{7}))$ with the cohomology class c_1 on $\widehat{Z}(\sqrt{7})$ is given by*

$$c_1\cdot F_B(\widehat{Z}(\sqrt{7})) = 2\int \omega + \sum_x Z_x\cdot F_B,$$

where the integral is over $\Gamma_B\backslash\mathcal{H}^2$, the integrand is the volume form

$$\omega = -\frac{1}{2\pi}\frac{dx\wedge dy}{y^2},$$

the summation runs runs over all of the cusps x of $\widehat{Z}(\sqrt{7})$ and where for a cusp x, we denote by Z_x the sum of all of the curves S_i^x in the resolution cycle of the cusp x.

PROOF. See [van der Geer 1988, Cor. VII.4.1] and the explicit description of the local Chern cycle at x on pp. 46 and 63 of the same reference. The result is not stated in there in a way that makes it completely clear that it is also valid for congruence subgroups, but that is implicit in its proof. □

LEMMA 7.14. *The self-intersection number of $Z_N(\widehat{Z}(\sqrt{7}))$ for $N = 14, 28$ is*

$$Z_N(\widehat{Z}(\sqrt{7}))^2 = \begin{cases} -2 & \text{if } N = 14, \\ -6 & \text{if } N = 28. \end{cases}$$

PROOF. We already know that the arithmetic genus of $Z_N(\widehat{Z}(\sqrt{7}))$ for $N = 14, 28$ is equal to the volume of $Z_N(\widehat{Z}(\sqrt{7}))$. It therefore follows from the adjunction formula and the preceding lemma that the self-intersection number of $Z_N(\widehat{Z}(\sqrt{7}))$ is equal to its volume of $Z_N(\widehat{Z}(\sqrt{7}))$. We are therefore done by Corollary 4.2. □

We now verify that the curves $Z_N(\widehat{Z}(\sqrt{7}))$, for $N = 14, 28$, are hyperelliptic.

LEMMA 7.15. *Let $N = 14, 28$. Then the intersection number of $Z_N(\widehat{Z}(\sqrt{7}))$ with*

$$F_1(\widehat{Z}(\sqrt{7})) + F_2(\widehat{Z}(\sqrt{7})) + F_4(\widehat{Z}(\sqrt{7}))$$

is equal to 6 if $N = 14$ and is equal to 10 if $N = 28$. The curve $Z_N(\widehat{Z}(\sqrt{7}))$ is hyperelliptic, with the involution τ inducing the hyperelliptic involution on $Z_N(\widehat{Z}(\sqrt{7}))$.

PROOF. The computation of the intersection number follows at once from the table in Lemma 7.9. Since $F_N(\widehat{Z}(\sqrt{7}))$ has 21 components, an odd number, and since these components must be permuted by the involution τ of $\widehat{Z}(\sqrt{7})$, there must be at least 1 component invariant under τ. Since τ commutes with the action of $\mathrm{PSL}_2(\mathbb{F}_7)$, it follows that every component $Z_N(\widehat{Z}(\sqrt{7}))$ is invariant under τ. The number of fixed points of τ on $Z_N(\widehat{Z}(\sqrt{7}))$ is equal to the intersection number of $Z_N(\widehat{Z}(\sqrt{7}))$ with the fixed point set of τ, which by Lemma 7.7 is

$$F_1(\widehat{Z}(\sqrt{7})) + F_2(\widehat{Z}(\sqrt{7})) + F_4(\widehat{Z}(\sqrt{7})).$$

The intersection number computed in the first part of this lemma therefore gives the number of fixed points of τ on each component. Since $Z_N(\widehat{Z}(\sqrt{7}))$ has genus 2 and 6 fixed points for τ, when $N = 14$, and genus 4 and 10 fixed points for τ when $N = 28$, it follows that $Z_N(\widehat{Z}(\sqrt{7}))$ is a hyperelliptic curve and τ is its hyperelliptic involution. □

LEMMA 7.16. *The curve $F_N(\widehat{Z}(\sqrt{7}))$ has genus* 3 *for* $N = 1$ *and genus* 10 *for* $N = 2, 4$.

PROOF. By Corollary 4.2, the volume of $F_N(\widehat{\Gamma}(\sqrt{7}))$ is -28 for $N = 1$ and -42 for $N = 2, 4$. Therefore these numbers are also their Euler numbers. The curve $F_N(\widehat{Z}(\sqrt{7}))$ for $N = 1, 2, 4$ is obtained by adding the points where it meets the resolution cycles at the cusps. According to Lemma 7.7 (cf. Figure 2), there is one such point for each cusp, so there are 24 points in all. Therefore the Euler number of $F_N(\widehat{Z}(\sqrt{7}))$ is $-28 + 24 = -4$ for $N = 1$ and is $-42 + 24 = 18$ for $N = 2, 4$. Writing the Euler number as $2 - 2g$, we have $g = 3$ for $N = 1$ and $g = 10$ for $N = 2, 4$. □

8. The Symmetric Hilbert Modular Surface $W = \widehat{Z}(\sqrt{7})/\tau$

We denote by W the orbit space for the action of the switching involution τ on $\widehat{Z}(\sqrt{7})$. It follows from Lemma 6.3 and Lemma 7.6 that the natural mapping of $\widehat{Z}(\sqrt{7})$ onto W is a two-sheeted covering branched along $F_1(W) \cup F_2(W) \cup F_4(W)$. Furthermore, since τ commutes with the elements of G acting on $\widehat{Z}(\sqrt{7}))$, the group G acts on W.

LEMMA 8.1. *The surface W is an algebraic surface whose Betti numbers are given by*:

$$b_i = \begin{cases} 1 & \textit{if } i = 0, 4, \\ 0 & \textit{if } i = 1, 3, \\ 94 & \textit{if } i = 2. \end{cases}$$

PROOF. First of all, a surface of Hilbert modular type is always an algebraic surface, by the results of Baily and Borel [1966] (cf. [van der Geer 1988, Prop. II.7.1, p. 44]). It follows that the surface W is also an algebraic surface. Furthermore, Hilbert modular surfaces have vanishing first Betti number [van der Geer 1988, comments following Cor. IV.6.2, p. 82]. So $b_1 = 0$ and therefore, by Poincaré duality, $b_3 = 0$. Since $\widehat{W}(\sqrt{7})$ is connected, we have $b_0 = 1$ and then $b_4 = 1$ by

Poincaré duality. Therefore, the Euler number $e(W)$ satisfies

$$e(W) = b_2 + 2,$$

so we will prove b_2 has the value we claim by proving that $e(W) = 96$. Since $\widehat{Z}(\sqrt{7})$ is a two-sheeted covering branched of W branched along $F_1(W) \cup F_2(W) \cup F_4(W)$, we have

$$e(\widehat{Z}(\sqrt{7})) = 2e(W) - e(F_1(\widehat{Z}(\sqrt{7}))) - e(F_2(\widehat{Z}(\sqrt{7}))) - e(F_4(\widehat{Z}(\sqrt{7}))).$$

By [van der Geer 1988, Theorem IV.1.2, p. 60], the Euler numbers of the open curves $F_N(\widehat{\Gamma}(\sqrt{7})\backslash\mathcal{H}^2)$, for $N = 1, 2, 4$, are equal to their volumes, which we have computed in Corollary 4.2. We obtain these open curves by deleting the 24 points where $F_N(\widehat{Z}(\sqrt{7}))$ meets the resolution cycles of the cusps, $N = 1, 2, 4$. Therefore

$$e(F_N(\widehat{Z}(\sqrt{7}))) = \begin{cases} -4 & \text{if } N = 1, \\ -18 & \text{if } N = 2, 4; \end{cases}$$

hence

$$e(\widehat{Z}(\sqrt{7})) = 2e(W) + 4 + 18 + 18 = 2e(W) + 40.$$

Therefore, we just have to prove that $e(\widehat{Z}(\sqrt{7})) = 232$. By [van der Geer 1988, Theorem IV.2.5, p. 64], we have

$$e(\widehat{Z}(\sqrt{7})) = \operatorname{vol}(\widehat{\Gamma}(\sqrt{7})\backslash\mathcal{H}^2) + 5\cdot 24,$$

since there are 5 components in each of the 24 resolution cycles. Since the Hilbert modular group for $k = \mathbb{Q}(\sqrt{7})$ has index 2 in the extended Hilbert modular group and since $\widehat{\Gamma}(\sqrt{7})$ has index 168 in the extended Hilbert modular group $\widehat{\Gamma}$, it follows from [van der Geer 1988, Theorem IV.1.1, p. 59] that

$$\operatorname{vol}(\widehat{\Gamma}(\sqrt{7})\backslash\mathcal{H}^2) = 168\zeta_k(-1),$$

where ζ_k denotes the Dedekind zeta function of the quadratic field k. Using [van der Geer 1988, Theorem I.6.5, p. 20], we have

$$\zeta_k(-1) = \frac{1}{60}\sum_{x\in\mathbb{Z}} \sigma_1\left(\frac{28 - x^2}{4}\right) = \frac{2}{3},$$

where $\sigma_1(n)$ is the sum of the divisors of n if n is a positive integer and otherwise is 0. Therefore

$$\operatorname{vol}(\widehat{\Gamma}(\sqrt{7})\backslash\mathcal{H}^2) = 168\cdot\frac{2}{3} = 112,$$

hence

$$e(\widehat{Z}(\sqrt{7})) = \operatorname{vol}(\widehat{\Gamma}(\sqrt{7})\backslash\mathcal{H}^2) + 5\cdot 24 = 232.$$

This completes the proof of the lemma. $\square$

We now compute the intersection numbers of the modular curves and their components on W.

LEMMA 8.2. *We have the following table of intersection numbers on W.*

	$F_1(W)$	$F_2(W)$	$F_4(W)$	$F_{14}(W)$	$F_{28}(W)$
$F_1(W)$	-56	0	0	0	84
$F_2(W)$	0	-36	0	42	84
$F_4(W)$	0	0	-84	84	42
$F_{14}(W)$	0	42	84	-21	84
$F_{28}(W)$	84	84	42	84	105

PROOF. The intersection product of two divisors on W is half the intersection product of their pullbacks to $\widehat{Z}(\sqrt{7})$. Since the curves $F_N(W)$ for $N = 1, 2, 4$ form the branch locus of the natural mapping of $\widehat{Z}(\sqrt{7})$ onto W, we have

$$F_M(W)\cdot F_N(W) = c_{MN} F_M(\widehat{Z}(\sqrt{7}))\cdot F_N(\widehat{Z}(\sqrt{7})),$$

where

$$c_{MN} = \begin{cases} 2 & \text{if } M, N = 1, 2, 4, \\ 1 & \text{if } M = 1, 2, 4 \text{ and } N = 14, 28, \\ 1 & \text{if } M = 14, 28 \text{ and } N = 1, 2, 4, \\ \frac{1}{2} & \text{if } M, N = 14, 28. \end{cases}$$

Using this, the lemma follows at once from Lemma 7.8. We leave it to the reader to verify the computation. □

LEMMA 8.3. *Each component of $F_{14}(W)$ has self-intersection number -1. Each component of $F_{28}(W)$ has self-intersection number -3. For $N = 14, 28$, each component meets exactly 4 components of $F_{42-N}(W)$. Each such intersection is a transverse at a single point. Each component $F_N(W)$ is a rational curve for $N = 14, 28$.*

PROOF. Since the components of $F_N(\widehat{Z}(\sqrt{7})$ are not ramified in the covering $\widehat{Z}(\sqrt{7}) \to W$, the intersection numbers of these components with themselves, with each other and with components of $F_{42-N}(\widehat{Z}(\sqrt{7}))$ are one half of the corresponding intersection numbers for components of $F_N(\widehat{X}(\sqrt{7}))$ and $F_{42-N}(\widehat{X}(\sqrt{7}))$. Therefore we are done by the preceding lemma. □

The next lemma describes the image of the resolution cycles of $\widehat{Z}(\sqrt{7})$ in W.

LEMMA 8.4. *The resolution cycle Z_x on $\widehat{Z}(\sqrt{7})$ of a cusp x is mapped to a reducible curve with 3 components $\mathcal{C}_0^x$, $\mathcal{C}_1^x$, $\mathcal{C}_4^x$. For s modulo 5, the image of the component C_s^x is the curve $\mathcal{C}_t^x$ with $t = s^2$. The intersection numbers of the these*

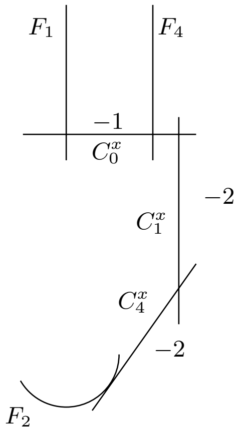


Figure 3. How the curves $F_N(W)$, for $N = 1, 2, 4$, meet the image of a resolution cycle in W.

curves are given by

$$\mathcal{C}_i \cdot \mathcal{C}_j = \begin{cases} -1 & \text{if } i = j = 0, \\ -2 & \text{if } i = j \neq 0, \\ 0 & \text{if } i = 0 \text{ and } j = 4, \\ 1 & \text{otherwise.} \end{cases}$$

Proof. The first assertion follows from the fact (arising from Lemma 7.6 and its proof) that τ interchanges the components C_s^x and C_{-s}^x. As for the intersection numbers, we have

$$\begin{aligned}
\mathcal{C}_0^x \cdot \mathcal{C}_0^x &= \tfrac{1}{2} S_0^x \cdot S_0^x = \tfrac{1}{2}(-2) = -1, \\
\mathcal{C}_0^x \cdot \mathcal{C}_1^x &= \tfrac{1}{2} S_0^x \cdot (S_1^x + S_4^x) = \tfrac{1}{2}(1+1) = 1, \\
\mathcal{C}_0^x \cdot \mathcal{C}_4^x &= \tfrac{1}{2} S_0^x \cdot (S_2^x + S_3^x) = \tfrac{1}{2}(0+0) = 0, \\
\mathcal{C}_1^x \cdot \mathcal{C}_1^x &= \tfrac{1}{2}(S_1^x + S_4^x) \cdot (S_1^x + S_4^x) = \tfrac{1}{2}(-2-2) = -2, \\
\mathcal{C}_1^x \cdot \mathcal{C}_4^x &= \tfrac{1}{2}(S_1^x + S_4^x) \cdot (S_2^x + S_3^x) = \tfrac{1}{2}(2) = 1, \\
\mathcal{C}_4^x \cdot \mathcal{C}_4^x &= \tfrac{1}{2}(S_2^x + S_3^x) \cdot (S_2^x + S_3^x) = \tfrac{1}{2}(-3-3+2) = -2,
\end{aligned}$$

and we are done. □

Figure 3 summarizes the intersection properties of the curves $\mathcal{C}_i^x$ and the curves $F_N(W)$, for $N = 1, 2, 4$.

Proposition 8.5. *The curve F_{28}, the curves $\mathcal{C}_i^x$ and the components of the curve F_{14} represent a basis for the second homology group $H_2(W; \mathbb{C})$.*

Proof. The number of curves is $1 + 21 + 3 \cdot 24 = 94$, which is the same as the second Betti number of W. We can blow down the components of $F_{14}(W)$ one at a time (they have self-intersection number -1 by Lemma 8.3). Independently of that, we can blow down the curves $\mathcal{C}_0^x$ one at a time. Having done so, the curves $\mathcal{C}_1^x$ are mapped to curves with self-intersection number -1, which can then be blown down one at a time. After that is done, the curves $\mathcal{C}_i^x$ are mapped to curves with self-intersection number -1, which can then be blown down one at a time. Each time one of these curves is blown down, the Betti number of W is decreased by 1 and when all have been blown down, the resulting surface has second Betti number $94 - 21 - 3\cdot 24 = 1$. This proves that the curves $\mathcal{B}_i^x$ and the components of F_{14} are linearly independent in the homology of W. Suppose we can write F_{28} as a linear combination of these curves. Since the curve F_{28} is invariant, components $\mathcal{B}_i^x$ and $Z_{14}(W)$ equivalent under the action of G have the same coefficient in this linear combination. Therefore we can write

$$F_{28}(W) = c_{14}F_{14}(W) + c_0 \sum_x \mathcal{C}_0^x + c_1 \sum_x \mathcal{C}_1^x + c_4 \sum_x \mathcal{C}_4^x,$$

where c_{14}, c_0, c_1, c_4 are rational numbers. However, if we intersect both sides with the cycle

$$F_{28}(W) + 4F_{14}(W),$$

we obtain a contradiction. Indeed, we have

$$F_{28}(W)\cdot(F_{28}(W) + 4F_{14}(W)) = 105 + 4\cdot 84 = 441$$

but since $F_{14}(W)$ and $F_{28}(W)$ are disjoint from the cuspidal components and

$$F_{14}(W)\cdot(F_{28}(W) + 4F_{14}(W)) = 84 + 4\cdot -21 = 0,$$

we have

$$(F_{28}(W) + 4F_{14}(W))\cdot(c_{14}F_{14}(W) + c_0 \sum_x \mathcal{C}_0^x + c_1 \sum_x \mathcal{C}_1^x + c_4 \sum_x \mathcal{C}_4^x) = 0. \quad \square$$

9. The Projective Plane as a Hilbert Modular Surface

As noted in the proof of Lemma 8.5, we can blow down all of the components of the curve $F_{14}(W)$ as well as all of the cycles $\mathcal{C}_0^x + \mathcal{C}_1^x + \mathcal{C}_4^x$. The resulting surface will be denoted $\mathbb{P}$. Since the components being blown down are permuted among themselves by the action of G on W, the group G also acts on $\mathbb{P}$. We will prove in this section that W is isomorphic to the complex projective plane.

First we need to recall a rationality criterion [van der Geer 1988, VII.2.2, p. 161]:

LEMMA 9.1. *If S is a nonsingular algebraic surface with $b_1 = 0$ and if S contains either two intersecting exceptional curves or an irreducible curve C with $C^2 \geq 0$ and $K\cdot C < 0$, then S is rational.*

In order to apply this formula, we will need to compute some intersection numbers on $\mathbb{P}$.

LEMMA 9.2. *We have the following intersection numbers on* $\mathbb{P}$.

	$F_1(\mathbb{P})$	$F_2(\mathbb{P})$	$F_4(\mathbb{P})$	$F_{28}(\mathbb{P})$
$F_1(\mathbb{P})$	16	48	72	84
$F_2(\mathbb{P})$	48	144	216	252
$F_4(\mathbb{P})$	72	216	324	378
$F_{28}(\mathbb{P})$	84	252	378	441

PROOF. To compute the self-intersection number of $F_1(\mathbb{P})$, we can use Mumford's definition of intersection numbers. After all, the surface on which one is computing intersections doesn't *have* to be singular for it to work. An easy computation shows that

$$F_1(W \to \mathbb{P}) = F_1(W) + \sum_x (3\mathcal{C}_0^x + 2\mathcal{C}_1^x + \mathcal{C}_4^x),$$

since its intersection with any of the curves that were blown down to make $\mathbb{P}$ is 0. Therefore the self-intersection number of $F_1(\mathbb{P})$ on $\mathbb{P}$ is equal to that of $F_1(W \to \mathbb{P})$, which is

$$\begin{aligned} F_1(W \to \mathbb{P})\cdot F_1(W \to \mathbb{P}) &= F_1(W) + \sum_x (3\mathcal{C}_0^x + 2\mathcal{C}_1^x + \mathcal{C}_4^x)\cdot F_1(W \to \mathbb{P}) \\ &= F_1(W)\cdot(F_1(W) + \sum_x (3\mathcal{C}_0^x + 2\mathcal{C}_1^x + \mathcal{C}_4^x)) \\ &= -56 + 24\cdot 3\cdot 1 = 16. \end{aligned}$$

Similarly, it is straightforward to verify that

$$\begin{aligned} F_2(W \to \mathbb{P}) &= F_2(W) + 2\cdot F_{14}(W) + \sum_x (2\mathcal{C}_0^x + 2\mathcal{C}_1^x + 2\mathcal{C}_4^x)), \\ F_4(W \to \mathbb{P}) &= F_4(W) + 4\cdot F_{14}(W) + \sum_x (3\mathcal{C}_0^x + 2\mathcal{C}_1^x + \mathcal{C}_4^x)), \\ F_{28}(W \to \mathbb{P}) &= F_{28}(W) + 4\cdot F_{14}(W). \end{aligned}$$

From this it is easy to complete the table using the table of Lemma 8.2. □

THEOREM 9.3. *The surface* $\mathbb{P}$ *is isomorphic to* $\mathbb{P}^2(\mathbb{C})$. *In particular*, W *is a rational surface*.

PROOF. Denote by K the canonical class of $\mathbb{P}$. By the adjunction formula and the fact that $c_1 = -K$, we have

$$K\cdot F_1(\mathbb{P}) = -e(F_1(\mathbb{P})) - F_1(\mathbb{P})\cdot F_1(\mathbb{P}) = -(-4) - 4.4 = -12 < 0.$$

On the other hand, we already know that $F_1\cdot F_1 = 16 > 0$, so by the rationality criterion, it follows that $\mathbb{P}$ is rational. Since $b_2(\mathbb{P}) = 1$ and the self-intersection number of $F_1(\mathbb{P})$ is > 0, the same will be true for the self-intersection number of any curve on $\mathbb{P}$. In particular, $\mathbb{P}$ has no exceptional curves and is therefore a minimal model. Since $b_2(\mathbb{P}) = 1$, we can find a rational number c such that K is homologous to cF_1. Then we have

$$-12 = K \cdot F_1(\mathbb{P}) = cF_1(\mathbb{P})^2 = 16c,$$

so $c = -3/4$. Therefore we have

$$K^2 = c^2 F_1(\mathbb{P})^2 = \frac{9}{16} \cdot 16 = 9.$$

It now follows from the classification of minimal models of algebraic surfaces [van der Geer 1988, § VII.2, p. 160] that $\mathbb{P}$ is actually $\mathbb{P}^2(\mathbb{C})$. The rationality of W now follows from the fact that it is birationally equivalent to $\mathbb{P}$. □

COROLLARY 9.4. *The curves $F_N(P)$ for $N = 1, 2, 4, 28$ are plane curves of degrees* 4,12,18,21 *respectively.*

PROOF. This follows at once from Bezout's theorem and from the table of intersection numbers for these curves on $\mathbb{P}$. □

COROLLARY 9.5. *The action of G on $\mathbb{P}$ arises from a irreducible complex linear representation of degree* 3.

PROOF. The reader is referred to [Conway et al. 1985] for the character table and group of Schur multipliers of G, which provides the basis for the following argument. Any projective representation of degree 3 of G arises from a linear representation of $\mathrm{SL}_2(\mathbb{F}_7)$. Since the action of G on $\mathbb{P}$ is nontrivial and since $\mathrm{SL}_2(\mathbb{F}_7)$ has no nontrivial representation of degree < 3, the representation is irreducible. Furthermore, the irreducible linear representations of degree 3 of $\mathrm{SL}_2(\mathbb{F}_7)$ arise from linear representations of G, so we are done. □

COROLLARY 9.6. *The locus $F_{14}(\mathbb{P})$ is a* 21*-point orbit for G acting on $\mathbb{P}$. The curves $F_N(\mathbb{P})$ for $N = 1, 2, 4, 28$ are defined by polynomials invariant under the linear* 3*-dimensional representation of G. The locus $F_{28}(\mathbb{P})$ is the union of* 21 *lines permuted transitively by G.*

PROOF. The first assertion follows from the fact that G acts transitively on the components of $F_{14}(\widehat{Z}(\sqrt{7}))$ and the corresponding facts on W and $\mathbb{P}$. The last assertion follows from the fact that $F_{28}(\mathbb{P})$ has 21 components permuted transitively by G and the fact that it is a plane curve of degree 21. Finally, suppose $f = 0$ is a polynomial defining a curve in $\mathbb{P}$ invariant under G. Then for all $g \in G$, the polynomial f is mapped to a multiple of itself by g, say to $c_g f$. The function $g \mapsto c_g$ is then easily seen to be a homomorphism from G to the multiplicative group of $\mathbb{C}^\times$. Since G is a simple group, that character is trivial, which proves the second assertion. □

Now that we have identified $\mathbb{P}$, we can identify the surfaces W, $\widehat{Z}(\sqrt{7})$ and $\widehat{X}(\sqrt{7})$. The orbits in P mentioned in Lemma 9.7 and Lemma 9.8 are discussed in more detail in Section 11. (We thank Igor Dolgachev, who pointed out an error in the original statement of Lemma 9.7 and explained how to correct it.)

Lemma 9.7. *The surface* $\widehat{Z}(\sqrt{7})$ *is obtained from the complex projective plane through the following steps*:

(1) *Blow up the unique* 21*-point orbit* O_{21} *and the unique* 24*-point orbit* O_{24} *for* G *acting on* $\mathbb{P}$. *Call the resulting surface* P'.
(2) *For each point* x *of* O_{24}, *let* $E(x)$ *denote the exceptional curve in* P' *obtained by blowing up* x. *Let* F_1' *denote the proper transform of the Klein curve* $x^3y+y^3z+z^3x=0$ *in* P' *and for each* x *of* O_{24} *let* x' *be the point of* $E(x)$ *where* F_1' *meets* $E(x)$. *The points* x' *form a* 24*-point orbit* $\overline{O}_{24}$ *in* P'. *We blow up this orbit and call the resulting surface* P''.
(3) *For each point* x' *of* $\overline{O}_{24}$, *denote by* $E'(x')$ *the line in* P'' *obtained by blowing up* x'. *Let* F_1'' *denote the proper transform of* F_1' *in* P'' *and for each point* x' *of* $\overline{O}_{24}$ *denote by* x'' *the point where* F_1'' *meets* $E'(x')$ *Then the points* x'' *form a* 24*-point orbit* $\overline{\overline{O}}_{24}$ *on* P''. *Blow up the orbit* $\overline{\overline{O}}_{24}$ *and call the resulting surface* P'''. *The surface* P''' *is* G *equivariantly isomorphic to* W.
(4) *Denote by* $\mathcal{D}_N$ *the proper transform of* $F_N(\mathbb{P})$ *in* P''' *for* $N=1,2,4$. *Let* $\mathcal{Z}$ *denote the two-sheeted cover* P''' *branched along* $\mathcal{D}_1\cup\mathcal{D}_2\cup\mathcal{D}_4$. *Then* $\mathcal{Z}$ *is equivariantly isomorphic to* $\widehat{Z}(\sqrt{7})$.
(5) *The preimage in* $\mathcal{Z}$ *of each point* x *of* $\mathcal{O}_{24}$ *is a cycle* S^x *of curves as shown in Figure* 2. *The cycle* S^x *can be blown down to a double point. Blowing down all of the cycles* S^x *results in a surface* $\mathcal{X}$ *equivariantly isomorphic to* $\widehat{X}(\sqrt{7})$.

What prevents Lemma 9.7 from containing a complete characterization of the surfaces $\widehat{X}(\sqrt{7})$ and $\widehat{Z}(\sqrt{7})$ is that we have not yet identified the plane curves $F_N(\mathbb{P})$ for $N=1,2,4$. The rest of this article is devoted to the solution of this problem. The following lemma will be of fundamental importance for that purpose.

Lemma 9.8. *The curve* $F_4(\mathbb{P})$ *has singularities of order* ≥ 4 *on the* 21*-point orbit* O_{21}. *The curve* $F_2(\mathbb{P})$ *is singular along the* 21*-point orbit* O_{24} *and the* 24*-point orbit* O_{24}.

Proof. Since each component $Z_{14}(W)$ of $F_{14}(W)$ and each curve $\mathcal{C}_i^x$ is blown down to a point under the natural mapping of W onto $\mathbb{P}$, it will suffice to verify the following intersection numbers on W:

$$Z_{14}\cdot F_2(W)=2,\quad Z_{14}\cdot F_4(W)=4,\quad \mathcal{C}_2^x\cdot F_2(W)=2.$$

The first and second of these follow immediately from the table in Lemma 8.2 if one notes that multiplying these intersection numbers by 21 must give the corresponding intersection numbers for F_{14}. As for the last, the left side must equal the intersection number on $\widehat{Z}(\sqrt{7})$ given by

$$(S_2^x+S_3^x)\cdot F_2(\widehat{Z}(\sqrt{7}))=2. \qquad \square$$

10. The Ring of Invariants of G on $\mathbb{C}^3$

In this section, we recall some of the classical results from [Klein and Fricke 1890–92, vol. 1, § III.7, pp. 732 ff.] on the ring of invariants for a three-dimensional irreducible complex representation ρ of G. An account of some of these results may also be found in [Weber 1896, §§ 122–124]. However, since some of our computations depend essentially on the precise forms for these invariants, we have also computed them ourselves using the algebra program REDUCE 3.4 on a personal computer.

The first invariant is the invariant f of degree 4 given by

$$f = x^3y + y^3z + z^3x. \tag{10.1}$$

The curve $f = 0$ is denoted $\mathcal{C}$ and is referred to as the *Klein curve*. The next invariant, of degree 6, is denoted ∇ and is given, up to a constant factor, by the determinant of the matrix of second partials of f. Explicitly,

$$\nabla = \frac{1}{54}\begin{vmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} & \dfrac{\partial^2 f}{\partial x \partial z} \\[2ex] \dfrac{\partial^2 f}{\partial y \partial x} & \dfrac{\partial^2 f}{\partial y^2} & \dfrac{\partial^2 f}{\partial y \partial z} \\[2ex] \dfrac{\partial^2 f}{\partial z \partial x} & \dfrac{\partial^2 f}{\partial z \partial y} & \dfrac{\partial^2 f}{\partial z^2} \end{vmatrix} \tag{10.2}$$
$$= 5x^2y^2z^2 - xy^5 - yz^5 - zx^5 = 5x^2y^2z^2 - \sigma(xy^5),$$

where in general we write $\sigma(x^ay^bz^c)$ to denote $x^ay^bz^c + x^by^cz^a + x^cy^az^b$. We will refer to the curve $\nabla = 0$ as the *Hessian* of $\mathcal{C}$ and denote it $\mathcal{H}$. It is the locus of all points in the plane whose polar conics with respect to the Klein curve are line pairs.

The next invariant, of degree 14, is, up to a constant factor, the determinant of the 4×4 symmetric matrix obtained by bordering the matrix of second partials of f with the first partials of ∇. Denoting the matrix by C, we have

$$C = \frac{1}{9}\begin{vmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} & \dfrac{\partial^2 f}{\partial x \partial z} & \dfrac{\partial \nabla}{\partial x} \\[2ex] \dfrac{\partial^2 f}{\partial y \partial x} & \dfrac{\partial^2 f}{\partial y^2} & \dfrac{\partial^2 f}{\partial y \partial z} & \dfrac{\partial \nabla}{\partial y} \\[2ex] \dfrac{\partial^2 f}{\partial z \partial x} & \dfrac{\partial^2 f}{\partial z \partial y} & \dfrac{\partial^2 f}{\partial z^2} & \dfrac{\partial \nabla}{\partial z} \\[2ex] \dfrac{\partial \nabla}{\partial x} & \dfrac{\partial \nabla}{\partial y} & \dfrac{\partial \nabla}{\partial z} & 0 \end{vmatrix}$$
$$= \sigma(x^{14} - 34x^{11}y^2z - 250x^9yz^4 + 375x^8y^4z^2 + 18x^7y^7 + 126x^6y^3z^5).$$

(Klein incorrectly gives the coefficient -126 as 126. The invariant is written correctly in Weber's presentation.) We will denote the curve $C = 0$ by Σ. It is the locus of all points in the plane whose polar lines with respect to $\mathcal{H}$ are tangent to their polar conics with respect to $\mathcal{C}$.

Finally, there is the invariant K of degree 21, which is, up to a constant factor, the functional determinant of f, ∇ and C:

$$
\begin{aligned}
K = \frac{1}{14}\begin{vmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} & \dfrac{\partial f}{\partial z} \\ \dfrac{\partial \nabla}{\partial x} & \dfrac{\partial \nabla}{\partial y} & \dfrac{\partial \nabla}{\partial z} \\ \dfrac{\partial C}{\partial x} & \dfrac{\partial C}{\partial y} & \dfrac{\partial C}{\partial z} \end{vmatrix}
&= \sigma(x^{21} - 7x^{18}y^2z + 217x^{16}yz^4 - 308x^{15}y^4z^2 - 57x^{14}y^7 - 289x^{14}z^7 + 4018x^{13}y^3z^5 \\
&\quad + 637x^{12}y^6z^3 + 1638x^{11}y^9z - 6279x^{11}y^2z^8 + 7007x^{10}y^5z^6 - 10010x^9y^8z^4) \\
&\quad + 10296x^7y^7z^7.
\end{aligned}
$$

(Weber doesn't give the full expression for this invariant; instead he cites [Gordan 1880] and [Klein and Fricke 1890–92]. In the former work, p. 372, our K is denoted by Ω and is listed with the wrong coefficient 3472 for $x^7y^7z^7$; this error also occurs in the latter work, p. 734. It is easy to guess its origin: since a notation Σ similar to our σ was used, the term $3432x^7y^7z^7$ appearing inside a σ would equal $10296x^7y^7z^7$.) We will denote by Λ the curve $K = 0$. It is the locus of all points in the plane whose polar lines with respect to the curves $\mathcal{C}$, $\mathcal{H}$ and Σ are concurrent.

Klein's generators f, ∇, C, K of the ring of invariants are connected by the relation

$$
\begin{aligned}
K^2 = C^3 - 88C^2f^2\nabla - 256Cf^7 + 1088Cf^4\nabla^2 + 1008Cf\nabla^4 \\
+1728\nabla^7 - 60032f^3\nabla^5 + 22016f^6\nabla^3 - 2048f^9\nabla.
\end{aligned}
$$

The generating function of the ring of invariants is given by

$$\sum_{n=0}^{\infty} a_n t^n = \frac{t^{21}}{(1-t^4)(1-t^6)(1-t^{14})},$$

where a_n is the dimension of the space of G-invariant forms of degree n. The series begins

$$1 + t^4 + t^6 + t^8 + t^{10} + 2t^{12} + 2t^{14} + 2t^{16} + 3t^{18} + 2t^{20} + t^{21} + \dots \tag{10.3}$$

11. Orbits of G Acting on $\mathbb{P}^2(\mathbb{C})$

In this section, we determine the orbit structure of the group G acting on $\mathbb{P}^2(\mathbb{C})$. Our approach is based on the enumeration of the subgroups of G and the decomposition of ρ restricted to each subgroup. For that purpose, it is useful to have a list of all of the conjugacy classes of subgroups of G. That information is provided in the following proposition, due to Klein [1879, § 1] (as corrected in his collected works). In [Adler $\geq$ 1998], we will make a more systematic and general study of orbits. A discussion of the orbits of G on $\mathbb{P}^2(\mathbb{C})$ can also be found in [Weil 1974, §§ 116–121].

LEMMA 11.1. *The group G has subgroups of the following orders*: 1, 2, 3, 4, 6, 7, 8, 12, 21, 24, 168. *All are unique up to conjugacy except for*:

(1) *the groups of order* 24, *which form two conjugacy classes of subgroups interchanged by the outer automorphisms of G*;
(2) *the subgroups of order* 12 *which occur as normal subgroups of these groups of order* 24 *and which fall into two conjugacy classes*;
(3) *the three conjugacy classes of subgroups of order* 4, *which consist of a conjugacy class of cyclic subgroups of order* 4 *and two conjugacy classes of Klein* 4-*groups*.

G has no other subgroups.

For each element γ of G, the element $\rho(\gamma)$ has three eigenvalues, not necessarily distinct, forming a set with repetitions. That set with repetitions is denoted $\eta(\gamma)$.

In the next lemma and the following material, we follow the notation of [Conway et al. 1985] for the conjugacy classes.

LEMMA 11.2. *The eigenvalues of the elements of $\rho(G)$ are as follows*:

γ	*order*	$\eta(\gamma)$
$1A$	1	$\{1,1,1\}$
$2A$	2	$\{1,-1,-1\}$
$3A$	3	$\{1,\omega,\omega^2\}$
$4A$	4	$\{1,i,-i\}$
$7A$	7	$\{\zeta_7,\zeta_7^2,\zeta_7^4\}$
$7B$	7	$\{\zeta_7^3,\zeta_7^5,\zeta_7^6\}$

COROLLARY 11.3. *The identity element of G is the only element fixing all of $\mathbb{P}^2(\mathbb{C})$.*

Elements from the following conjugacy classes $3A$, $4A$, $7A$, *and* $7B$ *have only isolated fixed points on* $\mathbb{P}^2(\mathbb{C})$.

Elements from the remaining conjugacy class S^2 have a fixed line and a fixed point on $\mathbb{P}^2(\mathbb{C})$.

PROOF. This is obvious from Lemma 11.2. □

COROLLARY 11.4. *If a subgroup H of G has a fixed point in $\mathbb{P}^2(\mathbb{C})$, the order of H must be* 1, 2, 3, 4, 6, 7, *or* 8. *A subgroup of order* 6 *has a unique fixed point. A noncyclic subgroup of order* 4 *has* 3 *fixed points. A subgroup of order* 8 *has a unique fixed point. A subgroup of order* 4 *has* 3 *isolated fixed points. A subgroup of order* 7 *has* 3 *isolated fixed points. A subgroup of order* 2 *has a fixed line and a fixed point.*

PROOF. A subgroup H of G has a fixed point if and only if the three-dimensional representation of G restricts to a reducible representation of H. The cases where the restriction is irreducible can be found by a straightforward character computation. The remainder are as described in the statement of this corollary. To see that the fixed point is unique in the case of the subgroup of order 6, note that an element of order 2 normalizing an element $3A$ of order 3 sends $3A$ to its inverse and therefore interchanges the two fixed points of $3A$ corresponding to the eigenvalues ω, ω^2. It must leave the entire fixed point set of $3A$ invariant and therefore must fix the fixed point with eigenvalue 1. Therefore, a subgroup of order 6 has a unique fixed point. To see that a noncyclic subgroup H of order 4 has exactly 3 fixed points, note that an element a of order 2 has a fixed point corresponding to the eigenvalue 1 and a fixed line corresponding to the eigenvalue -1. If b is another element of order 2 of H, then b must leave the fixed point set of a invariant. Therefore it fixes the isolated fixed point of a and also has two fixed points of its own on the fixed line of a (it can't fix the fixed line of a since then ab would act as the identity). This proves the assertion about the noncyclic subgroups of order 4. Finally, a subgroup H of order 8 is the normalizer of a cyclic subgroup of order 4. Since an element of order 2 in H not lying in the cyclic subgroup of order 4 sends an element $4A$ of order 4 in H to its inverse, it will interchange the two fixed points of $4A$ corresponding to the eigenvalues $\pm i$. It will also necessarily fix the remaining fixed point of $4A$ corresponding to the eigenvalue 1. This proves the corollary. □

COROLLARY 11.5. *The group G has orbits of the following orders in its action of $\mathbb{P}^2(\mathbb{C})$:* 21, 24, 28, 42, 56, 84, 168. *The orbits of orders* 21, 24, 28, 42 *and* 56 *are unique. The* 42*-point orbit arises from the conjugacy class of cyclic subgroups of order* 4 *of G. There are ∞^1 orbits of order* 84 *and ∞^2 orbits of order* 168. *The closure of the union of the* 84*-point orbits coincides with the locus $K = 0$, where K is the invariant of degree* 21 *of Klein.*

PROOF. From the orders of the subgroups of G having fixed points, we know that the possible orders of orbits are 21, 24, 28, 42, 56, 84, 168. The uniqueness of the orbit of order 21 follows from the fact that a subgroup of order 8 is unique up to conjugacy and has a unique fixed point. A similar argument shows the uniqueness of the orbit of order 28. The uniqueness of the orbit of order 24 follows from that fact that an element $7A$ of order 7 is sent to $7A^2$ and $7A^4$ by the normalizer of the cyclic group it generates. Therefore, the normalizer acts transitively on the three fixed points of $7A$, which correspond to the eigenvalues $\zeta_7, \zeta_7^2, \zeta_7^4$. If H is a cyclic group of order 4, it has 3 fixed points. The fixed point belonging to the eigenvalue 1 is actually fixed by the normalizer of H, as we noted above, and therefore belongs to a 21-point orbit, not a 42-point orbit. The other two fixed points are interchanged by the normalizer of H and give rise to a single 42-point orbit. If H is a Klein 4-group then H has 3 fixed points. Each involution in H has a fixed line and an isolated fixed point. The isolated fixed points of the three involutions are fixed points of H as a whole. However, none of them can lie in a 42-point orbit since the isolated fixed point of an involution is, as we have already noted, in a 21-point orbit. Therefore, the Klein 4-groups contribute no 42-point orbits. That the union of the 84-point orbits has dimension 1 follows from the fact that an element a of order 2 has a fixed line and an isolated fixed point. Furthermore, that isolated fixed point lies on the fixed line of another element $b \neq a$ of order 2 provided b commutes with a. Therefore the union of the 84-point orbits is the same as the union of the fixed lines of the elements of order 2. Since there are 21 elements of order 2 in G, the union of their fixed lines will be the union of 21 lines and invariant under the action of G. The product of the linear forms defining these 21 lines will then be an invariant of degree 21 and must therefore coincide, up to a nonzero scalar factor, with Klein's invariant of degree 21. A cyclic subgroup H of order 3 has 3 fixed points, two of which are exchanged by the normalizer of H and which therefore give rise to a single orbit of order 56. The remaining fixed point is also a fixed point of the normalizer of H and gives rise to a 28-point orbit. Since the orbits of order < 168 consist of a finite number of points and a finite number of lines, there must remain ∞^2 orbits of order 168. □

NOTATION 11.6. We will denote by O_d an orbit of order d. When such an orbit is unique, there is no ambiguity in this notation. In order to resolve the ambiguity in the case of the orbits of orders 84 and 168, we can denote an orbit by $O_d(p)$ where p is a point of the orbit.

Our next goal is to give a list of explicit orbit representatives for the orbits of orders 21, 24, 28, 42 and 56. This computation was carried out using REDUCE 3.4 on a personal computer.

LEMMA 11.7. *For* $d = 21, 24, 28, 42$, *the point* p_d *given below is a representative of the orbit* O_d.

$$\begin{aligned} p_{21} = \big[&\sqrt{-7}(-\zeta_7^2-3\zeta_7-3) + 7(\zeta^2+\zeta_7+1),\\ &2\sqrt{-7}(\zeta_7^2-1),\\ &\zeta_7(\sqrt{-7}\zeta_7-\sqrt{-7}+7\zeta_7+7)\big] \end{aligned}$$

$$
\begin{aligned}
p_{24} &= [1,0,0]\\
p_{28} &= [1,1,1]\\
p_{42} &= \big[\sqrt{-7}(-i\zeta_7^2-4i\zeta_7+\zeta_7-1-2i) + 7(-1+\zeta_7-2i+i\zeta_7^2),\\
&\qquad 2\sqrt{-7}(i\zeta_7^2-i\zeta_7+\zeta_7-1),\\
&\qquad i\sqrt{-7}(\zeta_7^2+2\zeta_7+4)+\sqrt{-7}(-3\zeta_7-4) + 7\zeta(-i\zeta_7+2\zeta_7+1)\big]\\
p_{56} &= [1,\omega,\omega^2]
\end{aligned}
$$

PROOF. We will merely explain the method by which these values were obtained. The matrices $\rho(2A)$, $\rho(3A)$, $\rho(4A)$ and $\rho(7A)$ are given in an appendix to this article, where ρ is the three-dimensional representation of G we are considering. As noted above, one obtains a point of O_{21} by taking an eigenvector of $\rho(4A)$ corresponding to the eigenvalue 1. Since the eigenvalues of $\rho(4A)$ are $1, \pm i$, the operator $\rho(4A)^2+1$ maps $\mathbb{C}^3$ into the 1 eigenspace. So we can take

$$
p_{21} = (\rho(4A)^2+1)\begin{pmatrix}1\\0\\0\end{pmatrix},
$$

which leads to the value for p_{21} obtained above. The 24-point orbit consists of the fixed points of $\rho(7A)$, of which $[1,0,0]$ is clearly one, so we have p_{24}. The 28-point orbit consists of the points belong to the eigenvalue 1 of the elements of order 3 of G. Since cyclic permutation of the coordinates is such an element and since $[1,1,1]$ clearly arises from an eigenvector with eigenvalue 1, we have p_{28}. To obtain p_{42}, we need to find an eigenvector of $\rho(4A)$ with eigenvalue i. The operator $(\rho(4A)-1)(\rho(4A)+i)$ clearly projects onto the i-eigenspace and we obtain p_{42} by applying this operator to the point $[1,0,0]$. The orbit O_{56} arises from the ω- and ω^2-eigenvectors of $\rho(3A)$, a cyclic permutation of the coordinates. So we can take p_{56} to be $[1,\omega,\omega^2]$. □

12. Characterization of the Hessian of Klein's Quartic

Denote by $\mathcal{C}$ the plane curve $x^3y+y^3z+z^3x=0$ and by $\mathcal{H}$ its Hessian.

PROPOSITION 12.1. *The plane curve $\mathcal{H}$ is defined by*

$$
5x^2y^2z^2 - xy^5 - yz^5 - zx^5 = 0,
$$

is irreducible, nonsingular, has genus 10 *and admits G as a group of automorphisms.*

PROOF. This is the expression given in (10.1) for the invariant ∇ of degree 6 for G. Since G has no permutation representation of degree < 7, any factor of ∇ is also an invariant. We see from the generating function (10.3) that ∇ is therefore irreducible. For if ∇ were reducible, the only factor it could have would be the invariant of degree 4, whereas 6 is not a multiple of 4. If $\mathcal{H}$ were singular, its singular locus would contain a G-orbit. Since every orbit of G on $\mathcal{P}^2(\mathbb{C})$ has at least 21 points, $\mathcal{H}$ would have at least 21 points of the same multiplicity ≥ 2. By the Plücker formula, an irreducible sextic curve can't have more than 10 singular points. Therefore $\mathcal{H}$ is nonsingular and it follows from the Plücker formula that $\mathcal{H}$ ias genus 10. □

PROPOSITION 12.2. *Let* Γ *denote the group whose presentation is*

$$\langle a, b, c \mid a^2 = b^4 = c^7 = abc = 1\rangle.$$

Let ϕ *be any surjective homomorphism from* Γ *to* G*. Then there is an automorphism* α *of* G *such that* $\alpha \circ \phi(a)$, $\beta \circ \phi(b)$ *and* $\gamma \circ \phi(c)$ *are the elements of* G *represented respectively by*

$$\begin{pmatrix} 0 & 3 \\ 2 & 0 \end{pmatrix}, \quad \begin{pmatrix} 3 & 5 \\ 4 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

PROOF. We know that $\phi(c)$ must satisfy

$$\phi(c)^7 = 1.$$

If $\phi(c) = 1$ then

$$1 = \phi(abc) = \phi(a)\phi(b),$$

which would imply that the image of ϕ is commutative and that ϕ is not surjective. Therefore $\phi(c)$ has order 7. Since the group of automorphisms of G acts transitively on the set of elements of order 7 in G, we can assume without loss of generality that $\phi(c)$ is the element of G represented by $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. The same reasoning as before shows that $\phi(a)$ must have order 2. There are 21 elements of order 2 in G and they are all conjugate under the upper triangular subgroup of G. Therefore, $\phi(a)$ is of the form $g\left(\begin{smallmatrix} 0 & 3 \\ 2 & 0 \end{smallmatrix}\right)g^{-1}$, where g is an upper triangular matrix of determinant 1. We can write g in the form

$$g = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y & 0 \\ 0 & y^{-1} \end{pmatrix}.$$

Since $\left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right)$ commutes with $\phi(c)$, we may after composing ϕ with the inner automorphism determined by $\left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right)$ suppose without loss of generality that x is 0. Therefore,

$$\phi(a) = \begin{pmatrix} 0 & y^2 \\ y^{-2} & 0 \end{pmatrix}$$

and

$$\phi(b) = \phi(a^{-1}c^{-1}) = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & y^2 \\ y^{-2} & 0 \end{pmatrix} = \begin{pmatrix} -y^{-2} & y^2 \\ y^{-2} & 0 \end{pmatrix}.$$

If $\phi(b)^2 = 1$, then $\phi(b)$ has trace 0, which is impossible since the trace is $-y^{-2}$. Therefore, $\phi(b)$ has order 4 and its trace is ± 4, whence $y = \pm 3$. This proves the proposition. □

Corollary 12.3. *There is one and only one normal subgroup Δ of Γ such that Γ/Δ is isomorphic to G.*

Proof. Let $A = \left(\begin{smallmatrix} 0 & 3 \\ 2 & 0 \end{smallmatrix}\right)$, $B = \left(\begin{smallmatrix} 3 & 5 \\ 4 & 0 \end{smallmatrix}\right)$ and $C = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ viewed as elements of G. Then $A^2 = B^4 = C^7 = ABC = 1$, so there is a homomorphism $\phi_0 : \Gamma \to G$ such that $\phi_0(a) = A$, $\phi_0(b) = B$ and $\phi_0(c) = C$. Let Δ_0 denote the kernel of ϕ_0. Since A, B, C generate G, the group Γ/Δ_0 is isomorphic to G. This proves the existence. As for the uniqueness, let Δ be any normal subgroup of Γ such that Γ/Δ is isomorphic to G. Then there is a surjective homomorphism $\phi : \Gamma \to G$ whose kernel is Δ. By Proposition 12.2, there is an automorphism α of G such that $\alpha \circ \phi = \phi_0$. Consequently the kernel of ϕ equals the kernel of ϕ_0, so $\Delta = \Delta_0$. $\square$

Theorem 12.4. *There is one and up to isomorphism only one smooth curve of genus 10 whose automorphism group contains a group isomorphic to G. Any such curve is equivariantly isomorphic to the plane curve $\mathcal{H}$.*

Proof. The last assertion follows from the first combined with Proposition 12.1. The same proposition shows that such curves exist, so we only have to prove that any two such curves are isomorphic. Let S be a compact Riemann surface of genus 10 on which G acts nontrivially. Let S' denote the orbit space for this action of G on S. Then S' naturally has the structure of a compact Riemann surface. Let g denote the genus of S'. For every point $\mathcal{P}$ of S let $e_{\mathcal{P}}$ denote the order of the stabilizer of $\mathcal{P}$ in G. If γ is any element of G then the stabilizer of $\mathcal{P}$ is conjugate via γ to the stabilizer of $\gamma \cdot \mathcal{P}$. Therefore $e_{\mathcal{P}}$ depends only on the orbit of $\mathcal{P}$ under G. If p is the point of S' representing that orbit, we may also write e_p for $e_{\mathcal{P}}$.

Using the Riemann–Hurwitz relation between the Euler characteristic of S and that of S', we have

$$-18 = 2 - 2(10) = 168(2 - 2g) - \sum_{\mathcal{P} \in S} (e_{\mathcal{P}} - 1).$$

If we group together all the terms in the summation which belong to the same orbit and instead sum over S', we obtain

$$-18 = 168(2 - 2g) - \sum_{p \in S'} \frac{168}{e_p}(e_p - 1) = 168(2 - 2g) - 168 \sum_{p \in S'} \left(1 - \frac{1}{e_p}\right).$$

Dividing through by -168, we get

$$\frac{3}{28} = 2g - 2 + \sum_{p \in S'} \left(1 - \frac{1}{e_p}\right).$$

Since the left-hand side is not an integer, the set of $p \in S'$ such that $e_p > 1$ must be nonempty and, of course, finite. Denote by n the number of points p of S'

with $e_p > 1$ and call these points $p_1, \ldots, p_n$. Also, we will write e_i for e_p when $p = p_i$. Then we have

$$\frac{3}{28} = 2g - 2 + \sum_{i=1}^{n} \left(1 - \frac{1}{e_i}\right).$$

If $g \geq 1$, the right-hand side of this equation is at least $\frac{1}{2}$, and so greater than $\frac{3}{28}$. Therefore $g = 0$. Furthermore, 7 divides the denominator of the left-hand side, so one of the e_i must be a multiple of 7. However, the stabilizer of any point of S must be cyclic, so actually one of the e_i equals 7, say, $e_1 = 7$. Then we have

$$\frac{5}{4} = \sum_{i=2}^{n} \left(1 - \frac{1}{e_i}\right).$$

If $n > 3$, the right-hand side is at least $\frac{3}{2}$, and so greater than $\frac{5}{4}$; hence $n \leq 3$. If $n < 3$, the right-hand side is less than 1, and so less than $\frac{5}{4}$; hence $n = 3$. Therefore

$$\frac{1}{e_2} + \frac{1}{e_3} = \frac{3}{4}.$$

If e_2, e_3 are both > 2, the left-hand side is at most $\frac{2}{3}$, and so less than $\frac{3}{4}$; hence one of e_2, e_3 is 2, say, $e_2 = 2$, and then $e_3 = 4$.

This proves that S is isomorphic to a Galois covering of $\mathbb{P}^1(\mathbb{C})$ with Galois group G and branched at exactly 3 points, the orders of branching being 2,4 and 7. The three points may, after applying a suitable projective transformation of $\mathbb{P}^1(\mathbb{C})$, be taken to be 0, 1 and ∞ respectively. The fundamental group Π of $\mathbb{P}^1(\mathbb{C})$ with these 3 points removed may be presented as a free group on the letters a, b, c with the relation $abc = 1$. Any branched cover of $\mathbb{P}^1(\mathbb{C})$ branched only at $0, 1, \infty$ to orders 2,4,7 respectively corresponds uniquely to a subgroup Δ of the group Π modulo the relations $a^2 = b^4 = c^7 = abc = 1$, that is to say, of the group Γ defined in Propostion 12.2. Furthermore, if such a cover is Galois then Δ is a normal subgroup and the Galois group is isomorphic to Γ/Δ. In our case, let Δ correspond to the branched cover $S \to S' = \mathbb{P}^1(\mathbb{C})$, so that Γ/Δ is isomorphic to G. Then Δ is the kernel of a surjective homomorphism from Γ to G. By Corollary 12.3, there is one and only one normal subgroup Δ' of Γ such that Γ/Δ' is isomorphic to G. It folows that there is only one possibility for Δ and, since Δ determines the branched cover, only one possibility for S. This completes the proof of the theorem. □

Corollary 12.5. *Hirzebruch's curves F_2 and F_4 are birationally equivalent to the Hessian of Klein's quartic.*

Corollary 12.6. *The curve $\mathcal{H} : xy^5 + yz^5 + zx^5 - 5x^2y^2z^2$ arises from an arithmetic subgroup of* $\mathrm{SL}_2(\mathbb{R})$.

Proof. We know that $\mathcal{H}$ arises from a subgroup of the $\{2, 4, 7\}$ triangle group, which was shown to be arithmetic by Fricke [1893b]. □

Remark 12.7. The fact that $\mathcal{H}$ is the normalization of F_2 and F_4 and the explicit definition of these curves in terms of congruence subgroups of unit groups of rational quaternion algebras also implies that the curve $\mathcal{H}$ arises from arithmetic groups. However, by construction, the curves F_2 and F_4 also arise from congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$.

Remark 12.8. While browsing through old journals recently, we found the article [Fricke 1893a]. From it, we learned that not only did he know that the $\{2, 4, 7\}$ triangle group is arithmetic but he also anticipated most of the results of this section one century ago. Time and space do not allow a detailed discussion of Fricke's papers here, but we will return to them in [Adler $\geq$ 1998].

Remark 12.9. Dolgachev has kindly pointed out that Theorem 12.4 also follows from Corollary 14.7 and Lemma 14.1.

13. The Jacobian Variety of the Hessian

Lemma 13.1. *The Jacobian variety of $\mathcal{H}$ is G-equivariantly isogenous to the product of an abelian variety A of dimension 3 and an abelian variety B of dimension 7, both of which admit G as an automorphism group. Furthermore, A is isomorphic to the Jacobian variety of the Klein curve $\mathcal{C}$ and therefore to a product of 3 copies of the elliptic curve $\mathbb{C}/L$, where L is the ring of integers of the imaginary quadratic field $\mathbb{Q}(\sqrt{-7})$, and B is isogenous to the sum of 7 copies of the elliptic curve*

$$PQR = S^3, \quad P + Q + R = 5S.$$

Proof. The plane cubics give the adjoint system for $\mathcal{H}$. Explicitly, every holomorphic differential on $\mathcal{H}$ is obtained by taking the residue along $\mathcal{H}$ of a rational differential of the form

$$\frac{E}{5x^2y^2z^2 - xy^5 - yz^5 - zx^5}\left(\frac{dy}{y} \wedge \frac{dz}{z} + \frac{dz}{z} \wedge \frac{dz}{x} + \frac{dx}{x} \wedge \frac{dy}{y}\right),$$

where E runs over the space of all cubic forms in x, y, z. Consequently, the action of G on the space of holomorphic differentials of $\mathcal{H}$ may be identified with the action of G on the space of ternary cubics. Since the polars of points of $\mathbb{P}^2(\mathbb{C})$ with respect to $\mathcal{C}$ form a covariant system of cubics, it follows that there is a 3-dimensional irreducible space V of cubic forms generated by the partials of $x^3y + y^3z + z^3x$. The ten-dimensional representation is therefore the sum of an irreducible representation of degree 3 and a representation of degree 7. Since the degrees of the irreducible representations of G are 1, 3, 6, 7, 8, the seven-dimensional space W of cubics can be reducible only if it contains the trivial representation. Since there is no G-invariant cubic, it follows that the invariant seven-dimensional space of differentials is irreducible. This implies at once that the Jacobian variety $J(\mathcal{H})$ of $\mathcal{H}$ contains an abelian variety A of dimension 3 invariant under G and an abelian variety B of dimension 7 invariant under G.

It is well known that A must be the Jacobian of $\mathcal{C}$. In [Adler 1981] one can find a proof of this and of the fact that $J(\mathcal{C})$ is isomorphic to the product of three copies of $\mathbb{C}/L$. Next consider the mapping $\lambda : \mathcal{H} \to \mathbb{P}^3$ given by

$$\lambda([x, y, z]) = [xy^5, yz^5, zx^5, x^2y^2z^2] = [P, Q, R, S].$$

Then $\lambda \circ g = \lambda$, where g is the collineation

$$[x, y, z] \mapsto [\zeta x, \zeta^4 y, \zeta^2 z]$$

of $\mathbb{P}^2(\mathbb{C})$. The image of λ is the elliptic curve

$$PQR = S^3, \quad P + Q + R = 5D,$$

which we will denote $\mathcal{E}$. We therefore have a surjective mapping of $J(\mathcal{H})$ onto $J(\mathcal{E})$ invariant under g, and an elliptic curve $\mathcal{E}'$ in $J(\mathcal{H})$ which maps onto $J(\mathcal{E})$. Durthermore, g must leave $\mathcal{E}'$ pointwise fixed, so $\mathcal{E}' \subseteq B$. Since the action of G on the tangent space of B is irreducible, it follows that B is isogenous to the sum of 7 copies of $\mathcal{E}'$, or what is the same, of $\mathcal{E}$, since $\mathcal{E}'$ and $\mathcal{E}$ are isogenous. $\square$

It is natural to wonder whether $J(\mathcal{H})$ is actually G-equivariantly isomorphic to $A \oplus B$, not merely isogenous. That will be investigated in a sequence of lemmas. If r is a prime number, we will denote by I_r the group of points of order r in $A \cap B$. Then I_r is invariant under G and is a vector space over $\mathbb{F}_r$, so I_r gives a modular representation of G. The representations of G on the fundamental groups $\pi_1(A)$ and $\pi_1(B)$ are integral representations of degrees 6 and 14 respectively. Their extensions to $\mathbb{C}$ decompose as

$$\pi_1(A) \otimes \mathbb{C} \cong V \oplus V^*, \quad \pi_1(B) \otimes \mathbb{C} \cong W \oplus W^* \cong W \oplus W,$$

since $W \cong W^*$. We can reduce the integral represenations on $\pi_1(A)$ and $\pi_1(B)$ modulo r and then obtain modular representations each of which has a submodule isomorphic to I_r. The set of irreducible representations of G over the algebraic closure $\bar{\mathbb{F}}_r$ which occur as composition factors of $\pi_1(A) \otimes \bar{\mathbb{F}}_r$ will be denoted $\mathcal{A}_r$, while the set of those occurring as composition factors of $\pi_1(B) \otimes \bar{\mathbb{F}}_r$ will be denoted $\mathcal{B}_r$. Denote by $\mathcal{C}_r$ the intersection of $\mathcal{A}_r$ and $\mathcal{B}_r$. The composition factors of G acting on $I_r \otimes \bar{\mathbb{F}}_r$ must lie in $\mathcal{C}_r$.

Lemma 13.2. *If r does not divide the order of G then $I_r = (0)$.*

Proof. If r does not divide the order of G, the representations V and W remain irreducible modulo r. Therefore the sets $\mathcal{A}_r$ and $\mathcal{B}_r$ are disjoint, which implies $\mathcal{C}_r$ is empty and $I_r \otimes \bar{\mathbb{F}}_r$ has no composition factors. Therefore $I_r = 0$. $\square$

Lemma 13.3. $I_7 = (0)$.

Proof. It is well known that that both the seven-dimensional representation W and the three-dimensional representations V, V^* remain irreducible modulo 7. Therefore, $\mathcal{C}_7$ is empty and $I_7 = (0)$. $\square$

LEMMA 13.4. $I_3 = (0)$.

PROOF. We will show that the irreducible representation of degree 7 remains irreducible modulo 3. Since 7 is prime, this will imply that the representation is absolutely irreducible and therefore that $\mathcal{C}_3$ is empty and $I_3 = (0)$. Since 3 is a primitive root modulo 7, G has no nontrivial representation of degree < 6 over $\mathbb{F}_3$. Therefore, since W is nontrivial modulo 3, it is either irreducible modulo 3 or else has composition factors of degrees 1 and 6. Since W is self-dual, the reduction of W, if reducible, would contain the trivial representation. An explicit model of the representation is given by the functions on $\mathbb{P}^1(\mathbb{F}_7)$ with values in $\mathbb{F}_3$ and whose sum over $\mathbb{P}^1(\mathbb{F}_7)$ is 0. Such a function cannot be G-invariant, so the representation of dimension 7 is irreducible modulo 3 and we are done. □

LEMMA 13.5. *We have* $\mathcal{B}_2 = \{V, V^*, 1\}$ *and* $\mathcal{A}_2 = \mathcal{C}_2 = \{V, V^*\}$.

PROOF. The representation V is irreducible in all characteristics. Therefore the second statement follows from the first. To prove the first, we use an explicit model for the seven-dimensional representation modulo 2, namely the space X of all functions on $\mathbb{P}^1(\mathbb{F}_7)$ with values in $\mathbb{F}_2$ and whose sum over $\mathbb{P}^1(\mathbb{F}_7)$ is 0. The constant functions lie in X and form an invariant subspace Y of dimension 1. We identify $\mathbb{F}_2$-valued functions with subsets of $\mathbb{P}^1(\mathbb{F}_7)$. Modulo constant functions, this means that every subset is identified with its complement. It is not difficult to show that Y is reducible. Indeed, the orbit of the subset $\{\infty, 3, 5, 6\}$ under G consists of 7 subsets up to complements and these together with the empty set form a three-dimensional subspace of Y. □

We summarize the results of these lemmas in the following corollary.

COROLLARY 13.6. *The kernel of the natural G-equivariant homomorphism of $A \oplus B$ onto $J(\mathcal{H})$ induced by addition is a finite 2-group. The elements of order 2 in that group form a group G-equivariantly isomorphic to I_2, which is of order 1, 8 or 64 and whose composition factors lie in the set consisting of the natural three-dimensional representation of $GL_3(\mathbb{F}_2)$ on $\mathbb{F}_2^3$ and its contragredient representation.*

REMARK 13.7. The discussion so far does not settle the question of the existence of nontrivial elements of order 2 fixed by G. We will show in Lemma 14.1 that there is a G-invariant element of order 2 in B. We also note that we have left open the precise determination of the group I_2 as well as the question of the existence of points of order 2^n in $A \cap B$ with $n \geq 2$. I am informed by Fred Diamond that one can use the methods of Ribet [1983] to determine intersections of invariant abelian subvarieties of the Jacobian varieties of modular curves. It seems reasonable to expect that these methods could also adapt to the case of curves arising from arithmetic groups with compact quotient. Since the curve $\mathcal{H}$ is such a curve, as well as a curve arising from arithmetic subgroups of finite index in $\mathrm{SL}_2(\mathbb{Z})$, we can perhaps expect a precise determination of $A \cap B$ from these methods.

For later use, we also need the following result.

LEMMA 13.8. *Let X be a curve of genus* 10 *which admits G as an automorphism group. Then any equivariant rational mapping of X onto another curve on which G acts is either birational or else maps X onto a rational curve on which G acts trivially.*

PROOF. By Theorem 12.4 and Lemma 13.1, the representation of G on the holomorphic differentials of X is the sum of an irreducible representation of degree 3 and an irreducible representation of degree 7. It follows that if Y is a smooth irreducible complete curve on which G acts and if $\phi : X \to Y$ is a G-equivariant morphism then the genus g of Y can only be 0, 3, 7 or 10. If $g = 10$ then f is necessarily birational. If $g = 0$ then G necessarily acts trivially on Y. It remains to show that the cases $g = 3, 7$ don't occur.

Suppose $g = 3$. It is well known (see [Hecke 1935], for example) that Y must be isomorphic to the Klein curve. Let p be a point of X fixed by an element γ of order 4 of G. Then $\phi(p)$ must be a point of Y which is also fixed by γ. However, an element of order 4 has no fixed points on the Klein curve. So we cannot have $g = 3$.

To show that $g \neq 7$, it suffices to observe that in general there can be no nonconstant mapping of a curve of genus 10 onto one of genus 7. Indeed, the degree n of such a mapping would have to be at least 2 and by the Riemann–Hurwitz relation we would then have

$$-18 = 2 - 2{\cdot}10 \leq n{\cdot}(2 - 2{\cdot}7) \leq -24,$$

which is impossible. (I am indebted to Noam Elkies for this observation, which greatly simplified the argument.) $\square$

14. Invariant Line Bundles on the Hessian

Denote by $\mathcal{L}$ the group of isomorphism classes of G-invariant line bundles on the Hessian curve $\mathcal{H}$ and denote by $\mathcal{L}_0$ the subgroup of $\mathcal{L}$ represented by invariant line bundles of degree 0. The quotient group $\mathcal{L}/\mathcal{L}_0$ is the group of all integers which occur as the degrees of G-invariant line bundles and will be denoted $\partial_{\mathcal{L}}$. Similarly, we denote by $\mathcal{M}$ the subgroup of $\mathcal{L}$ whose elements correspond to G-invariant divisors on the curve $\mathcal{H}$. We denote by $\mathcal{M}_0$ the subgroup of $\mathcal{M}$ consisting of elements of degree 0, so that

$$\mathcal{M}_0 = \mathcal{M} \cap \mathcal{L}_0.$$

The subgroup of $\partial_{\mathcal{L}}$ represented by elements of $\mathcal{M}$ is denoted $\partial_{\mathcal{M}}$.

LEMMA 14.1. *The group $\mathcal{M}_0$ is cyclic of order* 2 *and lies in the G-invariant abelian subvariety B of $J(\mathcal{H})$. In particular B (and a fortiori $J(\mathcal{H})$) has a G-invariant element of order* 2. *The group $\partial_{\mathcal{M}}$ is equal to* $6\mathbb{Z}$. *The group $\mathcal{M}$ is the product of $\mathcal{M}_0$ and an infinite cyclic group.*

PROOF. Denote by (p_2), (p_4) and (p_7) the G orbits on $\mathcal{H}$ of orders 84, 42 and 24 respectively. Identifying these orbits with the divisors they determine, it is easy to see, using the fact that the orbit space for G acting on $\mathcal{H}$ is $\mathbb{P}^1(\mathbb{C})$, that $2(p_2)$, $4(p_4)$ and $7(p_7)$ are linearly equivalent to each other and to any 168-point orbit. Therefore, $\mathcal{M}$ is generated by the line bundles associated to (p_2), (p_4) and (p_7). Those line bundles will be denoted ξ_2, ξ_4 and ξ_7 respectively. A line bundle represents an element of $\mathcal{M}_0$ if and only if it is of the form

$$\xi_2^a \xi_4^b \xi_7^c$$

where

$$84a + 54b + 24c = 0.$$

By solving this diophantine equation for a, b, c, we see that $\mathcal{M}_0$ is generated by $\xi_2\xi_4^{-2}$ and $\xi_4^4\xi_7^{-7}$. However, the line bundle $\xi_4^4\xi_7^{-7}$ is trivial since $7(p_7)$ and $4(p_4)$ are linearly equivalent. Therefore, $\mathcal{M}_0$ is a cyclic group generated by $\xi_2\xi_4^{-2}$. Since $2(p_2)$ is linearly equivalent to $4(p_4)$, the square of the line bundle $\xi_2\xi_4^{-2}$ is trivial. Therefore, $\mathcal{M}_0$ has order 1 or 2. Suppose the order is 1. Then (p_2) is linearly equivalent to $2(p_4)$. Let f be a rational function on $\mathcal{H}$ whose divisor is $2(p_4) - (p_2)$. Since the divisor $2(p_4) - (p_2)$ is G-invariant, the scalar multiples of f form a one-dimensional representation space for G. Since G is a simple group, that one-dimensional representation must be trivial. Therefore, f is fixed by every element of G and therefore is really a rational function on the orbit space $\mathbb{P}^1(\mathbb{C})$ for G acting on $\mathcal{H}$. In particular, the divisor of f is the preimage under the quotient mapping $\mathcal{H} \to \mathcal{H}/G = \mathbb{P}^1(\mathbb{C})$ of a divisor on $\mathbb{P}^1(\mathbb{C})$. Since the divisor is supported on the 42-point orbit and the 84-point orbit, it would then have to be of the form

$$2s(p_2) + 4t(p_4),$$

and that is a contradiction. Therefore $\mathcal{M}_0$ is cyclic of order 2. As for the group ∂_M, it is clearly generated by the greatest common divisor of the orders of the possible G-orbits on $\mathcal{H}$, i.e. of 24, 42, 84 and 168, which is 6. This proves the lemma. □

LEMMA 14.2. *Let H be a finite group acting on an algebraic curve X over the field $\mathbb{C}$ of complex numbers. Let $Y = X/H$ be the orbit space for the action of H on X and assume that Y is of genus 0. Denote by $\mathcal{L}_X$ the group of isomorphism classes of invariant line bundles on X and by $\mathcal{M}_X$ the subgroup of $\mathcal{L}_X$ consisting of elements represented by line bundles associated to H-invariant divisors. Then the quotient group $\mathcal{L}_X/\mathcal{M}_X$ is isomorphic to a subgroup of the group of Schur multipliers of H. In particular, the index of $\mathcal{M}_X$ in $\mathcal{L}_X$ divides the order of the group of Schur multipliers of H.*

PROOF. Denote by K_X the function field of X and by $K_X^\times$ the multiplicative group of K_X. Let $A = K_X^\times/\mathbb{C}^\times$, let B denote the group of all divisors on X and let C denote the group of all isomorphism classes of line bundles on X. Then we have the exact sequence

$$1 \to A \to B \to C \to 1.$$

From the associated long exact cohomology sequence of G, we have

$$1 \to A^H \to B^H \to C^H \to H^1(H, A).$$

From the exact sequence

$$1 \to \mathbb{C}^\times \to K^\times \to A \to 1,$$

we deduce that the sequence

$$H^1(H, K^\times) \to H^1(H, A) \to H^2(H, \mathbb{C}^\times)$$

is exact. However, $H^1(H, K^\times)$ is trivial by Hilbert's Theorem 90 and $H^2(H, \mathbb{C}^\times)$ is the group of Schur multipliers of H. Since $\mathcal{L}_X = C^H$ and since $\mathcal{M}_X$ is the image of B^H in $\mathcal{L}_X$, it follows that $\mathcal{L}_X/\mathcal{M}_X$ is isomorphic to a subgroup of the group of Schur multipliers of H. □

Originally, the following lemma merely asserted that $\mathcal{M}$ has index at most 2 in $\mathcal{L}$. I am greatly indebted to Dolgachev for communicating the proof of this stronger and more satisfactory form of the lemma.

Lemma 14.3. *The index of* $\mathcal{M}$ *in* $\mathcal{L}$ *is equal to* 2. *Furthermore, any torsion element of* $\mathcal{L}$ *lies in* $\mathcal{M}$. *In other words*, $\mathcal{L}_0 = \mathcal{M}_0$.

Proof. That the index is at most 2 follows at once from the preceding lemma and from the well known fact that the group of Schur multipliers of $\mathrm{PSL}(2,7)$ is cyclic of order 2. Dolgachev pointed out that in the preceding lemma the homomorphism $\mathrm{Pic}(X(p))^G \to H^2(G, \mathbb{C}^\times)$ is in fact surjective. We will present his argument in Lemma 14.4 below. Assuming this result, it follows that $\mathcal{L}/\mathcal{M}$ is a cyclic group of order 2. Let ξ be an invariant line bundle of degree 0 on $\mathcal{H}$. If ξ doesn't represent an element of $\mathcal{M}_0$ then the group $\mathrm{PSL}_2(\mathbb{F}_7)$ does not act on the bundle ξ. Instead the group $\mathrm{SL}_2(\mathbb{F}_7)$ will act and, in particular, the nontrivial element of the center of $\mathrm{SL}_2(\mathbb{F}_7)$ will act as -1 on the space of sections of ξ. On the other hand, since ξ is a nontrivial line bundle of degree 0, the Riemann–Roch theorem implies that the space of sections of ξ has dimension 9. However, all irreducible representations of $\mathrm{SL}_2(\mathbb{F}_7)$ in which the center acts nontrivially are of even dimension, which contradicts the fact that the space of sections must decompose into a direct sum of such representations. □

We now present Dolgachev's proof of the following lemma.

Lemma 14.4. *Let* X *be a curve and* $G \subseteq \mathrm{Aut}(X)$ *be a perfect group of automorphisms of* X. *Then the natural mapping of* $\mathrm{Pic}(X)^G$ *to* $H^2(G, \mathbb{C}^\times)$ *is surjective.*

Proof. The argument depends on two spectral sequences given in [Grothendieck 1957], namely

$$'E_2^{p,q} = H^p(G, H^q(X, \mathcal{O}_X^\times)) \Rightarrow H^{p+q}(G; X, \mathcal{O}_X^\times)$$
$$''E_2^{p,q} = H^p(G\backslash X, R^q\pi_*(\mathcal{O}_X^\times)) \Rightarrow H^{p+q}(G; X, \mathcal{O}_X^\times)$$

Here, as A runs over the category of G-sheaves of abelian groups, the functor $H^n(X; G, A)$ is the n-th right derived functor of the functor which associates to such a sheaf A its group of G-invariant sections. Also, π denotes the natural mapping of X onto $G\backslash X$. From these spectral sequences, one derives the following exact sequences [Grothendieck 1957, p. 201]:

$$0 \to H^1(Y, {\mathcal{O}_X^\times}^G) \to H^1(X; G, \mathcal{O}_X^\times) \to H^0(Y, H^1(G, \mathcal{O}_X^\times)) \to H^2(Y, {\mathcal{O}_X^\times}^G) \to H^2(X; G, \mathcal{O}_X^\times)$$
$$0 \to H^1(G, \mathbb{C}^\times) \to H^1(X; G, \mathcal{O}_X^\times) \to H^1(X, \mathcal{O}_X^\times)^G \to H^2(G, \mathbb{C}^\times) \to H^2(X; G, \mathcal{O}_X^\times)$$

The first spectral sequence has $E_2^{2,0} = E_2^{1,1} = E_2^{0,2} = 0$, which implies

$$H^2(G; X, \mathcal{O}_X^\times) = 0.$$

From this and the fact that $H^1(G, \mathbb{C}^\times) = 0$, we derive from the first exact sequence:

$$0 \to H^1(G; X, \mathcal{O}_X^\times) \to H^1(X, \mathcal{O}_X^\times)^G \to H^2(G, \mathbb{C}^\times) \to 1.$$

The group $H^1(X, \mathcal{O}_X^\times)^G$ is the group $\mathcal{L}$ and the group $H^1(G; X, \mathcal{O}_X^\times)$ is the group of line bundles on X with G-action. Such line bundles are precisely those arising from G-invariant divisors, so this group may be identified with the group $\mathcal{M}$. This proves the lemma. □

Corollary 14.5.

$$\partial_{\mathcal{L}} = 3\mathbb{Z}, \quad \mathcal{L}_0 = \mathbb{Z}/2\mathbb{Z}.$$

Proof. We know that $\mathcal{M}_0 = \mathbb{Z}/2\mathbb{Z}$ and $\partial_{\mathcal{M}} = 6\mathbb{Z}$. We also know that

$$\mathcal{L}/\mathcal{M} = \partial_{\mathcal{L}}/\partial_{\mathcal{M}} \oplus \mathcal{L}_0/\mathcal{M}_0 = \mathbb{Z}/2\mathbb{Z}.$$

We also know from the preceding lemma that $\mathcal{L}_0 = \mathcal{M}_0$. Therefore

$$\partial_{\mathcal{L}} = 3\mathbb{Z}, \quad \mathcal{L}_0 = \mathbb{Z}/2\mathbb{Z}.$$ □

Remark 14.6. It would be interesting to know something about the geometry of an embedding of $\mathcal{H}$ associated to a line bundle lying in $\mathcal{L}$ but not in $\mathcal{M}$. For example, by using contact quintics[1] of $\mathcal{H}$, one can map it to a curve of degree 15 in $\mathbb{P}^5$, where the action of $\mathrm{SL}_2(\mathbb{F}_7)$ on $\mathbb{P}^5$ is derived from an irreducible representation of degree 6 of $\mathrm{SL}_2(\mathbb{F}_7)$ in which the center acts nontrivially. It would probably not be difficult to write down the equations of the curve of degree 15 explicitly.

[1] A contact quintic of the sextic is one which is tangent to the sextic at every point where it meets the sextic. More precisely, it is a quintic which cuts out a divisor divisible by 2 in the group of all divisors on the sextic.

Corollary 14.7. *There are precisely two* G*-invariant curves of degree* 18 *and genus* 10 *in* $\mathbb{P}^2$.

Proof. Let $\mathcal{C}$ be such a curve and let $\phi : \mathcal{H} \to \mathcal{C}$ be the unique G-invariant mapping of $\mathcal{H}$ to $\mathcal{C}$. Let $\xi = \phi^*\mathcal{O}(1)$, where we denote by $\mathcal{O}(1)$ here the restriction to $\mathcal{C}$ of the ample generator of the Picard group of $\mathbb{P}^2(\mathbb{C})$. Then ξ is an invariant line bundle of degree 18 on $\mathcal{H}$. It follows from Corollary 14.5 and its proof that there are precisely two possibilities for ξ: either $\xi = K$ or $\xi = K \otimes \xi_0$, where K is the canonical line bundle of $\mathcal{H}$ and ξ_0 is the unique G-invariant line bundle of order 2 and degree 0. By the Wood's Hole Fixed Point Formula, we can write the Euler character of G on the cohomology of ξ as

$$\chi(\gamma;\xi) = \operatorname{tr}(\gamma \mid H^0(\xi)) - \operatorname{tr}(\gamma \mid H^1(\xi)) \sum \frac{\operatorname{tr}(\gamma \mid \xi_x)}{1 - d\gamma_x}$$

for γ in G, where the summation on the right runs over the fixed points x of γ on $\mathcal{H}$ and where $d\gamma_x$ and $\operatorname{tr}(\gamma \mid \xi_x)$ denote the scalar by which γ acts on the fibres K_x, ξ_x of K, ξ respectively at x. If $\xi = K$, we already know this Euler character and its decomposition from the proof of Lemma 13.1. It is $\chi_3' + \chi_7 - 1$. Now, the line bundle $\xi_0 = \xi_2\xi_4^{-2}$ is associated to the G-invariant divisor $(p_2) - 2(p_4)$. Therefore the scalar $\operatorname{tr}(\gamma \mid \xi_x)$ at a fixed point x of γ will be the same for both K and $K\xi_0$ except when γ has order i and x belongs to (p_i), with $i = 2, 4$. Leaving these traces undetermined for the moment, let $e_1 = \chi(\gamma; K\xi_0)$ where γ has order 4 and let $e_2 = \chi(\gamma; K\xi_0)$ where γ has order 2. For the remaining values we have $\chi(\gamma; K) = \chi(\gamma; K\xi_0)$. Therefore

$$K(\gamma; K\xi_0) = \begin{cases} 9 & \text{if } \gamma = 1A, \\ e_2 & \text{if } \gamma = 2A, \\ 0 & \text{if } \gamma = 3A, \\ e_1 & \text{if } \gamma = 4A, \\ \frac{1}{2}(-3 + \sqrt{-7}) & \text{if } \gamma = 7A, \\ \frac{1}{2}(-3 - \sqrt{-7}) & \text{if } \gamma = 7B. \end{cases}$$

On the other hand, we know from the proof of Lemma 14.3 that $H^1(K\xi_0) = 0$, so the Euler character of $K\xi_0$ is actually the character of G on sections of $K\xi_0$. Since an element of order 2 of G has at most 4 fixed points on $\mathcal{H}$, it follows that $|e_2| \le 2$. The multiplicity of χ_6 in the Euler character is then $(e_2 + 3)/4$. Since the multiplicity must be an integer, we must have $e_2 = 1$ and the multiplicity of χ_6 is 1. Similarly, we have $|e_1| \le 2$ and the multiplicity of χ_3' is $(e_1 + 3)/4$. Therefore, $e_1 = 1$ and the multiplicity of χ_3' is 1. Hence, the Euler character of $K\xi_0$ decomposes as $\chi_3' + \chi_6$. The mapping ϕ determines an embedding ϕ_x of the representation space of χ_3' into $H^0(\xi)$. Since χ_3' occurs with multiplicity 1 in both $H^0(K)$ and $H^0(K\xi_0)$, the embedding ϕ_ξ is uniquely determined up to a scalar factor for each line bundle ξ. Converely, ϕ_ξ determines the mapping ϕ and its image $\mathcal{C}$. This proves the lemma. $\square$

Remark 14.8. The techniques of [Dolgachev $\geq$ 1998] apparently do not apply directly to the study of G-invariant vector bundles on $\mathcal{H}$ since $\mathcal{H}$ is associated to a triangle group (2,4,7) and 2,4 are not relatively prime. But it is reasonable to hope that Dolgachev's methods can be extended to handle this case.

15. Identification of the Curve F_2 of Degree 12

The *Hessian* $\mathcal{H}$ is the locus of points in the plane whose polar conics with respect to the Klein curve $\mathcal{C}$ are pairs of lines. These lines are always distinct since $\mathcal{H}$ is nonsingular. The *Steinerian* of $\mathcal{C}$ is the locus of the point where these two lines meet. Denote the Steinerian by $\mathcal{S}$. There is a morphism ι from $\mathcal{H}$ to $\mathcal{S}$ which associates to the point p of $\mathcal{H}$ the singular point $\iota(p)$ of the polar conic of p with respect to $\mathcal{C}$.

The following result is due to Fricke [1893a, p. 386, eq. (3)]. It has been rediscovered in modern times by Dolgachev and Kanev [1993, p. 256, Ex. 6.1.1] and independently by the author.

Lemma 15.1. *The degree of* $\mathcal{S}$ *is* 12 *and its equation is*

$$4f^3 + \nabla^2 = 0. \tag{15.2}$$

Furthermore, S *has* 45 *double points. These consist of the* 24*-point orbit and the* 21*-point orbit*, *each taken once. The points of the* 24*-point orbit are cusps and those of the* 21*-point orbit are nodes.*

Proof. The graph of ι consists of all pairs $(p,q) \in \mathbb{P}^2(\mathbb{C}) \times \mathbb{P}^2(\mathbb{C})$ such that $M_p \cdot q = 0$, where M_p is the matrix of second partials of f at p. Consequently the graph is defined by 3 bihomogeneous equations of bidegree $(2,1)$. Denote by h the generator of the cohomology group $H^2(\mathbb{P}^2(\mathbb{C}), \mathbb{Z})$ and for $i = 1, 2$ denote by h_i the pullback of h via the projection π_i of $\mathbb{P}^2(\mathbb{C}) \times \mathbb{P}^2(\mathbb{C})$ onto its i-th factor. Then the graph of ι is Poincaré dual to

$$(2H_1 + h_2)^3 = 12h_1^2 h_2 + 6h_1 h_2^2.$$

The Steinerian $\mathcal{S}$ is simply the projection of the graph of ι onto the second factor and is Poincaré dual, at least as a cycle in $\mathbb{P}^2(\mathbb{C})$, to

$$(\pi_2)_*(2H_1 + h_2)^3 = 12h_2,$$

so $\mathcal{S}$ has degree ≤ 12. Since $\mathcal{S}$ is an irreducible invariant curve, we know from Klein's determination of the ring of invariants that $\mathcal{S}$ is either of degree 12 or else is either $\mathcal{H}$ or $\mathcal{C}$. By Lemma 13.7, $\mathcal{H}$ doesn't map onto $\mathcal{C}$, so $\mathcal{S} \neq \mathcal{C}$. If $\mathcal{S} = \mathcal{H}$, then ι would be a G-equivariant automorphism of $\mathcal{H}$, hence the identity. But it is easy to check that

$$\iota([1,0,0]) \neq [1,0,0].$$

This proves that $\mathcal{S}$ has degree 12. The general invariant of degree 12 is of the form

$$af^3 + b\nabla^2.$$

Since S has genus 10, it must have 45 multiple points, counted with their multiplicities. Since the 24-point orbit is the intersection of $\mathcal{C}$ and $\mathcal{H}$, it is clear that these are double points. The remaining 21 points form the 21-point orbit to which Hirzebruch refers [1977, p. 319]. Since neither $\mathcal{C}$ nor $\mathcal{H}$ passes through the 21-point orbit, there is only one invariant curve of degree 12 pasing through the 21-point orbit and that is $\mathcal{S}$. One of the 21 points is $[-y-z, y, z]$, where y and z are given by

$$y = \frac{\gamma^3 - \gamma^4}{\sqrt{-7}}, \quad z = \frac{\gamma^6 - \gamma}{\sqrt{-7}}. \tag{15.3}$$

By requiring that $\mathcal{S}$ pass through this point, we find that $\mathcal{S}$ is given by

$$4f^3 + \nabla^2 = 0.$$

To see that the 21-point orbit is double on $\mathcal{S}$, note that each cyclic group of order 4 in G has 3 fixed points, 2 of which lie on $\mathcal{H}$. The mapping ι sends the two on $\mathcal{H}$ to the remaining fixed point, which creates a double point on $\mathcal{S}$. More precisely, it has a node there since there are two points of the normalization $\mathcal{H}$ corresponding to it. As for the 24-point orbit, it is clear from the expression $4f^3 + \nabla^2$ that there is only one tangent at a point of the 24-point orbit, namely the tangent to $\mathcal{H}$ at that point, since $4f^3$ vanishes to third order at the point and ∇^2 only to second order. $\square$

16. Identification of the Curve F_4 of Degree 18

In Corollary 14.7, we showed that there are exactly two irreducible G-invariant plane curves of degree 18 and genus 10. In this section, we will describe them in more detail. As a first step in studying such curves, we note that there are 3 linearly independent invariants of degree 18 for G, namely fC, ∇^3 and $f^3\nabla$. Accordingly, there is a net of G-invariant curves of degree 18. We will denote that net by $\mathcal{N}$. We then have the following technical result, where we adopt the notation of § 11 for orbits.

LEMMA 16.1. *Denote by $\mathcal{N}$ the net of invariant curves of degree 18. Every element of the net $\mathcal{N}$ passes through the 24-point orbit and the 42-point orbit. If O_d is an orbit with d elements, then for $d = 21, 28, 56$ the elements of $\mathcal{N}$ passing through O_d form a pencil which we denote $\mathcal{P}_d$ and these three pencils are not concurrent in $\mathcal{N}$. For $d = 21, 28$, the orbit O_d is singular on any element of $\mathcal{N}$ containing it. The pencil $\mathcal{P}_{56}$ consists entirely of reducible curves. The elements of $\mathcal{N}$ singular on the orbit O_d form a pencil $\mathcal{P}_d$ for $d = 24, 42$. The pencil $\mathcal{P}_{24}$ consists entirely of reducible curves containing $\mathcal{H}$ as a component. There is one and only one element of $\mathcal{N}$ having multiplicity at least 3 at the points of O_{28}.*

The pencil $\mathcal{P}_{21}$ has a unique element for which the multiplicity at the points of O_{21} is at least 3 *and for that element the multiplicity is actually equal to* 4.

Proof. These assertions were all verified using REDUCE 3.4 on a personal computer. Some of them are easy to verify by hand. □

To show the existence of two essentially different types of invariant curves of degree 18 and genus 10, we begin with the following result which is based on a classical result for generic quartics [Berzolari 1903–15].

Proposition 16.2. *There is an irreducible invariant of degree* 18 *and genus* 10 *for G with all of the points of* 21*-point orbit as quadruple points and no other singularities. There is another irreducible invariant of degree* 18 *and genus* 10 *for G whose singularities are the points of the* 42*-point orbit and the points of an* 84*-point orbit. The former is the reflex* (*see below*) *of the Caylean of the Klein curve and the latter is the reflex of the Steinerian of the Klein curve.*

Proof. According to [Berzolari 1903–15, footnote 78 on p. 340 and table on p. 341], the dual of the Steinerian of a plane quartic has class 18, with 84 bitangents and 42 inflectional tangents. This means that the image of the Steinerian in $\mathbb{P}^2(\mathbb{C})^*$ (the dual $\mathbb{P}^2(\mathbb{C})$) under the Gauss map is a curve of degree 18 with 84 nodes and 42 cusps. Furthermore, in case the quartic is Klein's quartic, this image is invariant under G since S is. Now, the action of G on $\mathbb{P}^2(\mathbb{C})^*$ is related to the action of G on $\mathbb{P}^2(\mathbb{C})$ by an outer automorphism of G. Therefore, whenever an invariant curve U can be found in $\mathbb{P}^2(\mathbb{C})^*$ having certain properties, there will be a uniquely determined invariant curve, which we call the *reflex*[2] of U, in the original $\mathbb{P}^2(\mathbb{C})$ with the same properties as U. (Naturally, one cannot take this statement too literally. For example, the action of G on the reflex of U differs from the action of G on U by the outer automorphism. But the degree, the genus, the number of singularities, etc. will be the same for the curve and its reflex.) This observation allows us, in effect, to gloss over the distinction between invariant curves in $\mathbb{P}^2(\mathbb{C})$ and in $\mathbb{P}^2(\mathbb{C})^*$.

We know that the Steinerian of Klein's quartic is birationally equivalent to the Hessian, which has genus 10, and therefore any rational mapping from it onto another curve on which G acts nontrivially must be birational. In particular, the dual of the Steinerian will have genus 10 as well. Therefore the reflex of the dual of the Steinerian will be an invariant curve of degree 18 and genus 10 in $\mathbb{P}^2(\mathbb{C})$, which proves the second assertion of the proposition. As for the first, one works with the *Caylean* of a plane quartic. This is defined to be the curve in the dual $\mathbb{P}^2(\mathbb{C})$ consisting of of the lines joining points of the Hessian to the corresponding points of the Steinerian. The citation in [Be] also shows that the Caylean of a plane quartic has class 18 and has 21 quadruple tangent lines. In the case of Klein's quartic, one concludes that there is an invariant plane curve

[2] We don't refer to it as the dual of U because the dual of a curve already has a meaning which we will also be using.

of degree 18 in $\mathbb{P}^2(\mathbb{C})$ with 21 quadruple points, namely the *reflex of the Caylean* of Klein's quartic. Since the singularities are a union of orbits and the minimal G-orbit has 21 points, we obtain our 21 points which are at least quadruple. By the Plücker formula, we will know that the 21-point orbit is precisely quadruple and that there are no other singularities as soon as we show that the genus is 10. But by definition of the Caylean, there is a G equivariant rational mapping of the Hessian onto the Caylean which associates to a point x of the Hessian the line that joins x to the corresponding point of the Steinerian. This mapping must then be birational by Lemma 13.8.

The singular loci of invariant curves are unions of G orbits, but it is conceivable that there is some degeneration when specializing the general result to Klein's quartic. There can be no degeneration in the case of the reflex Caylean since there can be no orbit with less than 21 points. An argument is needed in the case of the reflex Steinerian, however, to verify that the singularities include the 42-point orbit and an 84-point orbit. We have verified this explicitly for the 42-point orbit but not for an 84-point orbit because of difficulties in computing the right 84-point orbit. The verification can however be completed by the following steps: (1) verify that the orbits O_{21}, O_{24}, O_{28}, O_{56} do not lie in the singular locus; (2) verify that the points of the 42-point orbit are nodes. For if (1) is verified, it will follow that all singularities of the reflex Steinerian lie on the 42-point orbit and on 84-point orbits. And if (2) is verified, it will follow that the 42-point orbit cannot be more singular than it is in the general case. In particular, O_{42} makes the same contribution to the Plücker formula as in the general case, so the same is true for the remaining orbits. Since the genus of the reflex Steinerian must be the same as the genus of $\mathcal{H}$, it follows that the remaining singularities must come from an 84-point orbit. The actual verification of (1) follows from Lemma 16.1. As for (2), the orbit O_{42} arises from i-eigenspaces in $\mathbb{C}^3$ of elements of order 4 in G. Since the eigenvalues of such an element are $1, \pm i$, the action of an element of order 4 on the tangent space to such a fixed point p has eigenvalues $-1, -i$ and acts, in suitable coordinates, by

$$(U, V) \mapsto (-U, -iV).$$

In the local ring at p, the defining equation must have the form

$$\phi(U, V) = 0,$$

where

$$\phi(U, V) = \sum_{n=0}^{\infty} \phi_n(U, V)$$

and where ϕ_n is the homogeneous part of degree n of ϕ. Since the point lies on the curve, we have $\phi_0 = 0$. Since the curve is defined by an invariant, we have

$$\phi_n(-U, -iV) = \phi_n(U, V)$$

for all n. This implies at once that $\phi_1 = 0$ and $\phi_2(U,V) = cU^2$. This proves that the points of the 42-point orbit are ordinary cusps of the curve, since we know that the curve can't have multiplicity greater than or equal to 3 on O_{42}. This completes the proof of Proposition 16.2. $\square$

REMARK 16.3. Another natural idea for getting another invariant curve of degree 18 and genus 10 is to consider the image of the Hessian under the dual mapping of the Klein curve, i.e. the mapping which associates to each point x of the Hessian the polar line of x with respect to the Klein curve. This mapping is certainly G-equivariant and the coordinates of the mapping are cubics, so one gets a curve of degree 18. The genus is again 10 for the same reasons as before. By Corollary 14.7, it must be one of the curves we have already mentioned. In fact, according to [Berzolari 1903–15], this curve is the same as the dual of the Steinerian curve. Berzolari refers to [Cremona 1861; Clebsch 1876; 1891; Kötter 1887; Voss 1887] for this beautiful result: *the line joining a point x of the Hessian to the corresponding point y of the Steinerian is tangent to the Steinerian at y!* We also note that since the cubics form the adjoint system for $\mathcal{H}$, the dual Steinerian can therefore be regarded as the projection of the canonical curve of $\mathcal{H}$ in $\mathbb{P}^9(\mathbb{C})$ from the unique G-invariant $\mathbb{P}^6(\mathbb{C})$ onto the unique G-invariant $\mathbb{P}^2(\mathbb{C})$.

COROLLARY 16.4. *The curve F_4 is the reflex of the Caylean of the Klein curve.*

PROOF. By Lemma 9.7, the curve F_4 does have quadruple points. The result now follows from Proposition 16.2. $\square$

Appendix: Matrices for Some Generators of G

We present here matrices according to which elements of each of the conjugacy classes of G act in the three-dimensional representation of G we are considering. The notation for representatives of conjugacy classes of G follows that of [Conway et al. 1985]. We do not claim that $4A^2 = 2A$, only that they be conjugate. One can also find a discussion of explicit matrices for this representation in [Weil 1974, § 115; Klein and Fricke 1890–92, § III.5, pp. 703–705].

$$7A = \begin{pmatrix} \zeta_7 & 0 & 0 \\ 0 & \zeta_7^4 & 0 \\ 0 & 0 & \zeta_7^2 \end{pmatrix} \qquad 7B = 7A^{-1} \qquad 3A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

$$2A = -\frac{1}{\sqrt{-7}} \begin{pmatrix} \zeta_7^2 - \zeta_7^5 & \zeta_7^4 - \zeta_7^3 & \zeta_7 - \zeta_7^6 \\ \zeta_7^4 - \zeta_7^3 & \zeta_7 - \zeta_7^6 & \zeta_7^2 - \zeta_7^5 \\ \zeta_7 - \zeta_7^6 & \zeta_7^2 - \zeta_7^5 & \zeta_7^4 - \zeta_7^3 \end{pmatrix} \qquad 4A = 7A^3 \cdot 2A$$

Acknowledgements

In addition to the obvious debt I owe to Hirzebruch, I also wish to thank Andrew Ogg for helpful discussions about Hilbert modular surfaces and for his extraordinary patience in listening to my often painful efforts to learn the subject. This phase of the work took place while I was at MSRI for 6 months in 1995–96; I'm grateful to Bill Thurston for inviting me to visit this unique institution and for considerable sensitivity in welcoming an independent scholar. I am also greatly indebted to Igor Dolgachev for reading the manuscript and pointing out a number of ways to improve the results. This includes, in particular, a sharpening of my inconclusive determination of the group of invariant line bundles on the Hessian of Klein's cubic. Knowing the precise group of invariant line bundles in turn made it possible to greatly shorten and simplify the arguments needed to identify Hirzebruch's curve F_4 of degree 18. Dolgachev also pointed out an error in the statement of Lemma 9.7 and explained how to correct it. Finally, I would like to thank Thorsten Ekedahl and Angelo Vistoli for their help with Lemmas 5.1 and 5.3.

References

[Adler 1979] A. Adler, "On Hirzebruch's curves F_1, F_2, F_4 for $\mathbb{Q}(\sqrt{7})$", preprint, 1979.

[Adler 1981] A. Adler, "Some integral representations of $\mathrm{PSL}_2(\mathbf{F}_p)$ and their applications", *J. Algebra* **72**:1 (1981), 115–145.

[Adler $\geq$ 2000] A. Adler, *On Hilbert modular surfaces and related topics*, In preparation.

[Ash et al. 1975] A. Ash, D. Mumford, M. Rapoport, and Y. Tai, *Smooth compactification of locally symmetric varieties*, Lie Groups: History, Frontiers and Applications **4**, Math. Sci. Press, Brookline, Mass., 1975.

[Baily and Borel 1966] J. Baily, W. L. and A. Borel, "Compactification of arithmetic quotients of bounded symmetric domains", *Ann. of Math.* (2) **84** (1966), 442–528. Reprinted as pp. 548–634 of Borel's *Oeuvres*, vol. 2, Springer, Berlin, 1983.

[Berzolari 1903–15] L. Berzolari, "Allgemeine theorie der höheren ebenen algebraischen Kurven", pp. 313–455 in *Encyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen*, Teubner, Leipzig, 1903–15.

[Borevich and Shafarevich 1966] A. I. Borevich and I. R. Shafarevich, *Number theory*, Pure and Applied Mathematics **20**, Academic Press, New York, 1966.

[Clebsch 1876] A. Clebsch, *Vorlesungen über Geometrie, 1: Geometrie der Ebene*, edited by F. Lindemann, Teubner, Leipzig, 1876. Reprinted by Johnson Reprint Corp., New York, 1968.

[Clebsch 1891] A. Clebsch, *Vorlesungen über Geometrie, 2: Theorie der Flachen erster und Zweiter Ordnung oder Klasse und der lineare Complex*, edited by F. Lindemann, Teubner, Leipzig, 1891. Reprinted by Johnson Reprint Corp., New York, 1968.

[Conway et al. 1985] J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, and R. A. Wilson, *Atlas of finite groups*, Oxford University Press, Oxford, 1985.

[Cremona 1861] L. Cremona, "Introduzione ad una teoria geometrica delle curve piane", *Memorie della R. Accademia delle scienze dell'Istituto di Bologna* (1) **12** (1861), 305–436. Annotated German translation by M. Curtze, Greifswald, 1865.

[Dolgachev $\geq$ 2000] I. Dolgachev, "Invariant stable bundles over modular curves", in *Recent Progress in Algebra* (Taejon, S. Korea, 1997), edited by H. C. Myung and E. Zelmanov, Contemporary Mathematics, Amer. Math. Soc. To appear.

[Dolgachev and Kanev 1993] I. Dolgachev and V. Kanev, "Polar covariants of plane cubics and quartics", *Adv. Math.* **98**:2 (1993), 216–301.

[Fricke 1893a] R. Fricke, "Entwicklungen zur Transformation fünfter und siebenter Ordnung einiger specieller automorpher Functionen", *Acta Mathematica* **17** (1893), 345–395.

[Fricke 1893b] R. Fricke, "Über den arithmetischen Charakter der zu den Verzweigungen $(2, 3, 7)$ und $(2, 4, 7)$ gehörenden Dreiecksfunctionen", *Math. Annalen* **41** (1893), 443–468.

[Fulton 1984] W. Fulton, *Intersection theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **2**, Springer, Berlin, 1984.

[van der Geer 1988] G. van der Geer, *Hilbert modular surfaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **16**, Springer, Berlin, 1988.

[Gordan 1880] P. Gordan, "Ueber die typische Darstellung der ternären biquadratischen Form $f = x_1^3x_2 + x_2^3x_3 + x_3^3x_1$", *Math. Annalen* **17** (1880), 359–378.

[Grothendieck 1957] A. Grothendieck, "Sur quelques points d'algèbre homologique", *Tōhoku Math. J.* (2) **9** (1957), 119–221.

[Grothendieck 1967] A. Grothendieck, "Éléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas (quatrième partie)", *Inst. Hautes Études Sci. Publ. Math.* **32** (1967), 1–361.

[Hausmann 1980] W. Hausmann, *Kurven auf Hilbertschen Modulflächen*, Bonner Mathematische Schriften **123**, Universität Bonn, 1980.

[Hecke 1935] E. Hecke, "Die eindeutige Bestimmung der Modulfunktionen q-ter Stufe durch algebraische Eigenschaften", *Math. Annalen* **111** (1935), 293–301.

[Hirzebruch 1971] F. E. P. Hirzebruch, "The Hilbert modular group, resolution of the singularities at the cusps and related problems", pp. 275–288 (exposé N° 396) in *Séminaire Bourbaki* (1970/1971), Lecture Notes in Math. **244**, Springer, Berlin, 1971. Reprinted as #48, pp. 192–205, of his *Gesammelte Abhandlungen*, vol. 2, Springer, 1987.

[Hirzebruch 1973] F. E. P. Hirzebruch, "Hilbert Modular Surfaces", *L'Énseignement Mathématique* **19** (1973), 183–282. Reprinted in the series Monographies de L'Enseignement Mathématique **21**, Univ. Genève, 1973.

[Hirzebruch 1977] F. E. P. Hirzebruch, "The ring of Hilbert modular forms for real quadratic fields in small discriminant", pp. 287–323 in *Modular functions of one variable, VI* (Bonn, 1976), edited by J.-P. Serre and D. B. Zagier, Lecture Notes in Math. **627**, Springer, Berlin, 1977.

[Hirzebruch 1979] F. E. P. Hirzebruch, 1979. Handwritten notes of computations.

[Hirzebruch 1995] F. E. P. Hirzebruch, Jul–Sep 1995. Private communication by email.

[Klein 1879] F. Klein, "Ueber die Transformationen siebenter Ordnung der elliptischen Funktionen", *Math. Annalen* **14** (1879), 428–471. Reprinted as [Klein 1923, LXXXIV, pp. 90–136]. Translated in this collection.

[Klein 1923] F. Klein, *Gesammelte Mathematische Abhandlungen, 3: Elliptische Funktionen* etc., edited by R. Fricke et al., Springer, Berlin, 1923. Reprinted by Springer, 1973.

[Klein and Fricke 1890–92] F. Klein, *Vorlesungen über die Theorie der elliptischen Modulfunctionen,* ausgearbeitet und vervollständigt von Robert Fricke (2 vol.), Teubner, Leipzig, 1890–92. Reprinted by Johnson Reprint, New York, 1966.

[Kötter 1887] E. Kötter, *Grundzuge einer rein geometrischen Theorie der algebraischen ebenen Curven*, Georg Reimer and Akademie der Wissenschaften, Berlin, 1887.

[Mumford 1961] D. Mumford, "The topology of normal singularities of an algebraic surface and a criterion for simplicity", *Publ. Math. Inst. Hautes Études Sci.* **9** (1961), 5–22.

[Ribet 1983] K. A. Ribet, "Mod p Hecke operators and congruences between modular forms", *Invent. Math.* **71**:1 (1983), 193–205.

[Satake 1956] I. Satake, "On a generalization of the notion of manifold", *Proc. Nat. Acad. Sci. U.S.A.* **42** (1956), 359–363.

[Satake 1960] I. Satake, "On compactifications of the quotient spaces for arithmetically defined discontinuous groups", *Ann. of Math.* (2) **72** (1960), 555–580.

[Voss 1887] A. Voss, "Zur Theorie der Hesse'schen Determinante", *Math. Annalen* **30** (1887), 418–424.

[Weber 1896] H. Weber, *Lehrbuch der Algebra*, Vieweg, Braunschweig, 1896. Reprinted by Chelsea, New York, 1961.

[Weil 1962] A. Weil, *Foundations of algebraic geometry*, 2nd ed., AMS Colloquium Publications **29**, Amer. Math. Soc., New York, 1962.

[Weil 1974] A. Weil, *Basic number theory*, Third ed., Grundlehren der Mathematischen Wissenschaften **144**, Springer, New York, 1974.

Allan Adler
P.O.Box 1043
Bowling Green, KY 42102-1043
United States
adler@hera.wku.edu

The Eightfold Way
MSRI Publications
Volume **35**, 1998

On the Order-Seven Transformation of Elliptic Functions

FELIX KLEIN
TRANSLATED BY SILVIO LEVY

This is a translation of Klein's *Ueber die Transformation siebenter Ordnung der elliptischen Funktionen*, first published in *Mathematische Annalen* **14** (1879), 428–471, and dated early November 1878. It follows the text printed in his *Gesammelte Mathematische Abhandlugen*, except where typos not present in the original had crept into formulas. I redrew all the figures (they had already been redrawn for the *Abhandlugen*: see caption on page 320), except for Figure 9 and the bottom figures on pages 315 and 316.

I have not attempted to modernize the terminology, except on a few occasions when the use of current language allowed me to replace a long-winded phrase by something crisper and clearer. Nor have I tried to approximate the English mathematical style of the time. The goal has been to produce a readable translation, as close to the original ideas as possible. Bibliographic citations have been converted to the house format, the editors of the *Abhandlugen* having taken similar liberties.

Brackets, if not delimiting bibliographic tags, indicate interpolated text, written either for the *Abhandlugen* (unsigned, or K. = Klein, B.-H. = Bessel-Hagen) or for this edition (L. = Levy).

I'm grateful to Jeremy J. Gray for many excellent suggestions.

In the study of the fifth-order transformation of elliptic functions we encounter, along with the modular equation of sixth degree and its well-known resolvent of fifth degree, the Galois resolvent of degree 60, called the *icosahedral equation*, which governs both. Starting from the icosahedral equation one sees with great ease the rule of formation and the properties of those lower-degree equations.

In this work I would like to further the theory of the transformation of the *seventh* order up to the same point. I have already shown in [Klein 1879a] how one can construct the modular equation of degree eight in its simplest form in terms of this theory. The corresponding resolvent of seventh degree was considered in [Klein 1879b]. The question now is *to construct the corresponding Galois resolvent of degree* 168 *in a suitable way, and to derive from it those lower-degree equations.*

As is well-known, the root η of this Galois resolvent, regarded as a function of the period ratio ω, has the characteristic property of remaining invariant under exactly those linear substitutions

$$\frac{\alpha\omega + \beta}{\gamma\omega + \delta}$$

that are congruent to the identity modulo 7. This will be for me in the sequel the definition of the irrationality η.

Therefore I begin (Section 1) with a short investigation of linear substitutions modulo 7. This investigation is thoroughly elementary, but should be included for the sake of completeness.[1] From this follows (Section 2) the way in which η is branched as a function of the absolute invariant J, and above all the fact that the equation linking η and J, which has genus $p = 3$, is sent to itself under 168 one-to-one transformations having an *a priori* specifiable arrangement.

This leads to a remarkable curve of order four, which is sent to itself under 168 collineations of the plane (Section 3) and which, as a consequence, enjoys a number of particularly simple properties (Sections 4 and 5). From the knowledge of the existence of those 168 collineations one can construct with little effort the whole system of covariants belonging to the curve (Section 6), and one obtains the equation of degree 168 in question in a particularly clear way, by intersecting the ground curve with a covariant pencil of curves of order 42 (still Section 6).

If one wants to descend from the equation so obtained to the modular equation of degree eight or to the resolvent of degree seven, certain results valid for the general curve of order four and dealing with contact curves of order three and with certain arrangements of bitangents (Sections 7–10) are particularly relevant. The roots of the equations under consideration thus turn out to be rational functions of the coordinates of *one* point on the curve, and to me the essential advance lies in this explicit representation achieved for the transformation of order seven.

The next several sections (Sections 11–15) attempt to sketch as intuitive as possible a picture of the branching of the Riemann surface defined by η as a function of J, and which is discussed more abstractly in Section 2. The figures that I have obtained in this way play the same important role in the understanding of the questions expounded here as the *shape* of the icosahedron plays in the related problem of degree five.

The most important results discussed here have already been announced in a note submitted on May 20, 1878 to the Erlangen Society.[2] There I had already shown how one can explicitly reduce those equations of degree seven having the same group as the modular equation to the modular equation itself.[3] In this article I do not yet go into this and other connected questions; I intend to return to them in more detail before long. [See [Klein 1879c].]

[1] Compare the more general investigations in [Serret 1866].

[2] [Klein 1878b]

[3] [In an earlier communication [Klein 1878a], I had only established the possibility of this reduction by abstract arguments. –K.]

1. Classification of the Substitutions $\dfrac{\alpha\omega+\beta}{\gamma\omega+\delta}$ Modulo 7

By a substitution $\dfrac{\alpha\omega+\beta}{\gamma\omega+\delta}$ pure and simple I will always mean a substitution

$$\omega' = \frac{\alpha\omega+\beta}{\gamma\omega+\delta}$$

where the coefficients are integers and the determinant is one. Moreover, for brevity, I will use the following expression. Two substitutions S_1 and S_2 are called *equivalent* if there is a third substitution S such that

$$S_1 = S^{-1} \cdot S_2 \cdot S.$$

In [Klein 1879a, § 8] I distinguished three kinds of such substitutions: elliptic, parabolic, and hyperbolic. The following propositions are straightforward:

Equivalent substitutions have the same sum $\alpha+\delta$.

All elliptic substitutions of period 2 (*and so satisfying* $\alpha+\delta=0$) *are equivalent.*

If elliptic substitutions of period 3 (*and so satisfying* $\alpha+\delta=\pm 1$) *are taken in pairs, so that one is the second iterate of the other, all such pairs are equivalent.*

Parabolic substitutions ($\alpha+\delta=\pm 2$) *fall into infinitely many classes, each contaning one representative among*

$$\omega' = \omega, \quad \omega' = \omega \pm 1, \quad \omega' = \omega \pm 2, \quad \ldots.$$

From now on we will consider substitutions $\dfrac{\alpha\omega+\beta}{\gamma\omega+\delta}$ only modulo 7, so we will regard two substitutions

$$\frac{\alpha\omega+\beta}{\gamma\omega+\delta} \quad \text{and} \quad \frac{\alpha'\omega+\beta'}{\gamma'\omega+\delta'}$$

as identical if $\alpha \equiv \alpha'$, $\beta \equiv \beta'$, $\gamma \equiv \gamma'$, $\delta \equiv \delta'$. Accordingly, we will not require that $\alpha\delta-\beta\gamma$ be equal to 1, but only that it be congruent to 1 modulo 7. In any case:

Substitutions that were formerly equivalent remain equivalent when considered modulo 7.

Now there are only finitely many substitutions, which can be easily counted:

The number of substitutions is 168.

Clearly, exactly one of these has period one, the identity $\omega' = \omega$. We will denote it by S_1.

To obtain the substitutions of period two, we introduce their characteristic condition, $\alpha+\delta=0$. There are 21 period-2 substitutions that are distinct modulo 7; since their period cannot change by considering them modulo 7, we have:

There are 21 *equivalent substitutions of period two*, which we denote by S_2. An example is $-1/\omega$.

In a similar way, applying the condition $\alpha+\delta=\pm 1$, which characterizes elliptic substitutions of period three, we obtain:

There are 28 *equivalent pairs of substitutions* S_3 *of period three*. An example of a pair is $-\frac{2}{3}\omega$, $-\frac{3}{2}\omega$.

For parabolic substitutions we had $\alpha+\delta=\pm 2$, which leads to 49 substitutions that are distinct modulo 7. One is the identity. Each of the others is equivalent to one of $\omega\pm 1$, $\omega\pm 2$, and $\omega\pm 3$, and so has period 7. Thus:

There are 48 *substitutions* S_7 *of period seven*, *divided into eight equivalent sextuples*. One sextuple would be $\omega+1$, $\omega+2$, …, $\omega+6$.

There remain $168-1-21-56-48=42$ substitutions, for which $\alpha+\delta=\pm 3$. The second iterate of a substitution of this kind satisfies $\alpha'+\delta'=0$, and so has period two; thus the 42 substitutions have period 4. I will pair each with its inverse. Thus:

There are 21 *equivalent pairs of substitutions* S_4 *of period four*, *each pair being associated with one* S_2.

For example, $\dfrac{2\omega+2}{-2\omega+2}$ and $\dfrac{2\omega-2}{2\omega+2}$ are associated with $-\dfrac{1}{\omega}$.

Connected with this classification of substitutions by their period is the construction of the *groups* they form. First we have the groups generated by a single element:

1. *One* G_1, *consisting of the identity only*: $\omega'=\omega$.
2. *Twenty-one* G_2*s with two substitutions each*; *for example* ω *and* $-1/\omega$.
3. *Twenty-eight* G_3*s with three substitutions each*; *for example* ω, $-\frac{2}{3}\omega$, $-\frac{3}{2}\omega$.
4. *Twenty-one* G_4*s with four substitutions each*; *for example*

$$\omega, \quad \frac{2\omega+2}{-2\omega+2}, \quad -\frac{1}{\omega}, \quad \frac{2\omega-2}{2\omega+2}.$$

5. *Eight* G_7*s with seven substitutions each*; *for example* ω, $\omega+1$, …, $\omega+6$.

Among these groups, any two that have the same number of elements are equivalent. For this reason, to prove each the following results, it is enough to exhibit one example satisfying the given description.

1. *Every* S_2 *commutes with exactly four other* S_2*s. These four fall into two pairs such that the elements of each pair commute with each other.*

Example: The substitution $-1/\omega$ commutes with

$$\frac{2\omega+3}{3\omega-2}, \quad \frac{3\omega-2}{-2\omega-3}, \quad \frac{2\omega-3}{-3\omega-2}, \quad \frac{3\omega+2}{2\omega-3}.$$

The first two of these commute, as do the last two.

2. *Thus there are* 14 *groups* G'_4 *of four elements such that every element different from the identity has period two*.[5] Examples:

$$\omega, \quad -\frac{1}{\omega}, \quad \frac{2\omega+3}{3\omega-2}, \quad \frac{3\omega-2}{-2\omega-3},$$

[5] [Such a group is called a *four-group* in my later terminology, which was picked up by other authors. –K.]

or

$$\omega, \quad -\frac{1}{\omega}, \quad \frac{2\omega-3}{-3\omega-2}, \quad \frac{3\omega+2}{2\omega-3}.$$

These 14 groups are *not* all equivalent; seven are equivalent to one of the examples just given, and seven to the other. Every G_2 is contained in a G_4' from each class.

3. *Every group G_3 commutes with exactly three S_2s. Thus there are* 28 *groups G_6' of six elements, all of them equivalent. Each S_2 is contained in four G_6's.* Example:

$$\omega, \quad -\frac{3\omega}{2}, \quad -\frac{2\omega}{3}, \quad -\frac{1}{\omega}, \quad \frac{2}{3\omega}, \quad \frac{3}{2\omega}.$$

4. *The four substitutions S_2s that, by item* 1 *above, commute with a given S_2 also commute with the G_4 that contains the given S_2. This gives* 21 *equivalent groups G_8' containing eight elements.*

Example: the four substitutions listed under item 4 on page 290, together with the four listed under item 1 on page 290.

5. *Each group G_7 commutes with* 14 *S_3s. This gives eight equivalent groups G_{21}' with* 21 *elements. Each S_3 lies in two of them.*

Example: $\omega+k$, $-\frac{2}{3}(\omega+k)$, $-\frac{3}{2}(\omega+k)$, for $k=0,1,\dots,6$; or again the set of all substitutions of the form

$$\frac{\alpha\omega}{\gamma\omega+\delta}.$$

6. *The* $2\cdot 7$ *groups G_4' (see item* 2 *above) give rise to* $2\cdot 7$ *groups G_{24}'' with* 24 *elements*, as follows. We take one G_4' and add to it:

(a) the six S_4s whose second iterates are in the chosen G_4';
(b) the six S_2s that commute with some S_2 from the G_4', but are not themselves in the G_4';
(c) the compositions of the six S_2s just mentioned, which together make four pairs of S_3s.

Adding up, we have $4+6+6+4\cdot 2=24$.

For example, take G_4' to consist of $\omega, -\dfrac{1}{\omega}, \dfrac{2\omega+3}{3\omega-2}, \dfrac{3\omega-2}{-2\omega-3}$. Then:

S_4s that belong to $-\dfrac{1}{\omega}$: $\dfrac{2\omega+2}{-2\omega+2}, \dfrac{2\omega-2}{2\omega+2}$.

S_4s that belong to $\dfrac{2\omega+3}{3\omega-2}$: $\dfrac{\omega+1}{\omega+2}, \dfrac{-2\omega+1}{\omega-1}$.

S_4s that belong to $\dfrac{3\omega-2}{-2\omega-3}$: $\dfrac{3\omega-3}{-3\omega+1}, \dfrac{\omega+3}{3\omega+3}$.

New S_2s that commute with $-\dfrac{1}{\omega}$: $\dfrac{2\omega-3}{-3\omega-2}, \dfrac{3\omega+2}{2\omega-3}$.

New S_2s that commute with $\dfrac{2\omega+3}{3\omega-2}$: $\dfrac{-\omega+1}{-2\omega+1}, \dfrac{\omega+2}{-\omega-1}$.

New S_2s that commute with $\dfrac{3\omega-2}{-2\omega-3}$: $\dfrac{3\omega-1}{3\omega-3}$, $\dfrac{-3\omega-3}{\omega+3}$.

Pairs of S_3s that arise by composition:

$$\frac{-3\omega-1}{2},\ \frac{-2\omega-1}{3};\quad \frac{2\omega}{\omega-3},\ \frac{3\omega}{\omega-2};\quad \frac{2}{3\omega+1},\ \frac{-\omega+2}{3\omega};\quad \frac{-\omega+3}{2\omega},\ \frac{-3}{2\omega+1}.$$

We see that the 24 substitutions making up a G''_{24} are related in the same way as the 24 permutations of four elements, or as the 24 rotations that take a regular octahedron to itself. I will make use later of both of these comparisons. These G''_{24} are obviously none other than the groups that I used in [Klein 1879b]. I wrote about these groups at that time in reference to Betti's work, in a slightly different form: namely, I did not stipulate that $\alpha\delta-\beta\gamma\equiv 1 \pmod 7$, but only that $\alpha\delta-\beta\gamma$ be congruent to a quadratic residue modulo 7—a distiction that has no significance in the context of fractional substitutions.

7. Finally, one can show by well-known methods that *the subgroups discussed above are the only ones to be found in the group of* 168 *substitutions in question.*[6]

2. The Function $\eta(\omega)$ and its Branching with Respect to J

Now let η be an algebraic function of J that is branched in such a way that, considered as a function of ω, it has the following properties:

1. It is single-valued everywhere in the positive half-plane ω.

2. It is sent to itself by exactly those substitutions $\dfrac{\alpha\omega+\beta}{\gamma\omega+\delta}$ that are congruent to the identity modulo 7.

I will denote by $\eta(\omega)$ one of the values corresponding to a given J. To obtain all other such values, it is enough to substitute for ω each of the 167 expressions $\dfrac{\alpha\omega+\beta}{\gamma\omega+\delta}$ that differ from ω modulo 7, because all the values of ω corresponding to the given J are of this form. It follows that:

η and J are related by an equation of degree 168 *in η.* We can denote the 168 roots (in some arbitrary order) by

$$\eta_1, \eta_2, \ldots, \eta_{168}.$$

Now the result of making J go around a closed path in the complex plane is to replace any of the associated ωs by $\dfrac{\alpha'\omega+\beta'}{\gamma'\omega+\delta'}$. Accordingly, the ηs undergo a certain permutation, as a result of the substitution of this fractional expression for ω in $\eta\left(\dfrac{\alpha\omega+\beta}{\gamma\omega+\delta}\right)$. If, after this permutation, *one* of the η_i coincides with its initial value (assuming J to be generic), the substitution $\begin{pmatrix}\alpha' & \beta'\\ \gamma' & \delta'\end{pmatrix}$ must be congruent to the identity modulo 7; therefore in this case *all* the η_i coincide with

[6] [An unsigned footnote in the *Abhandlungen* corrects this by mentioning the alternating group found within each symmetric group G''_{24}, and refers the reader to [Gierster 1881] –L.]

their initial values. For, if S denotes any substitution $\dfrac{\alpha\omega+\beta}{\gamma\omega+\delta}$ and S_0 any such substitution congruent to the identity modulo 7, the substitution S_0' such that

$$SS_0 = S_0'S$$

is likewise congruent to the identity modulo 7. Another way to express this is:

All roots η_i are equally branched with respect to J.[7]

The branch points themselves can only be at $J = 0, 1, \infty$, according to [Klein 1879a, § 2]. When J goes aroud 0, ω undergoes an elliptic substitution of period 2; when J goes around 1, ω undergoes an elliptic substitution of period 3; and when J goes around ∞, an appropriately chosen ω undergoes the parabolic transformation $\omega' = \omega + 1$. It follows that:

At $J = 0$ the 168 leaves of the Riemann surface that represents η are grouped into 56 cycles of three; at $J = 1$ they are grouped into 84 cycles of two; and at $J = \infty$, into 24 cycles of seven.

The *genus* of the equation that relates η and J is therefore found to be three:

$$p = \tfrac{1}{2}(2 - 2\cdot168 + 56\cdot2 + 84\cdot1 + 24\cdot6) = 3.$$

Two algebraic functions of J that have the same branching behavior are related by a rational expression in J. Thus:

Any root η_i of our equation is a rational function of any other root η_k and J.

Or, in other words:

One can construct 168 rational functions $R(\eta, J)$ with numerical coefficients, such that, if η is any of the roots, the others are given by

$$\eta_1 = R_1(\eta, J), \quad \eta_2 = R_2(\eta, J), \quad \ldots, \quad \eta_{168} = R_{168}(\eta, J).$$

Thus, corresponding to the 168 substitutions studied in Section 1, *there are 168 one-to-one transformations of our Riemann surface into itself*. The consequences that we are about to derive rest on the fact that we know the grouping of these substitutions from Section 1, and that these groups must have counterparts in terms of the one-to-one transformations of the Riemann surface that we will now consider.

We start with the following observation. The transformations take a point in the Riemann surface to another point directly above or below it [that is, one lying over the same J –L.]. If we ask, then, whether there are points that are left fixed by some transformations (or, equivalently, that are sent to less than 168 distinct images), the answer is simply the branch points: for they are the

[7] [The expression "equally branched" [gleichverzweigt] here means not merely that the branch locus of each root has the same arrangement and number of sheets as that of any other, but also that the *relationship* among the sheets for any two roots is the same. More precisely: If a point P describes on the (say) N-sheeted Riemann surface any simple closed curve, and if any other of the $N-1$ points exactly above or below P is made to move on the surface so as to always shadow P, it too will come back to itself. –B.-H.]

only points that belong to more than one sheet at once. Given what we said a moment ago about the branch locus, we conclude that:

Among the orbits of points arising from the group of transformations there is one with 24 *points of multiplicity seven*, *corresponding to* $J = \infty$; *one with* 56 *triple points*, *corresponding to* $J = 0$; *and one with* 84 *double points*, *corresponding* $J = 1$. *There are no other multiple points.*

I will give special names to these points, because of their importance: they will be called a-points, b-points, and c-points, respectively. Each a-point is fixed by a transformation of period 7, and thus by all the transformations of a G_7. Similarly, every b-point is fixed by a G_3, and every c-point by a G_2. But we know that there are only eight groups G_7, twenty-eight groups G_3 and twenty-one groups G_2, apart from twenty-one groups G_4. We conclude:

Each G_7 *leaves three a-points individually fixed*; *each* G_3 *leave two b-points fixed*; *and each* G_2 *leaves four c-points fixed.*

A G_4 *leaves no points fixed.*

Every G_7 was a normal subgroup of a G'_{21}, which apart from that contained only substitutions of period 3. The three a-points that are left fixed by a G_7 cannot be fixed by the other transformations in this G'_{21}; otherwise there would be only eight a-points in total, not 24. Therefore the three a-points are permuted by these other transformations, and since their period is three, the permutation is *cyclic*. Thus:

Every G'_{21} *has an associated triple of a-points that it leaves invariant as a set.*

In the same way one concludes:

Every G'_6 *has an associated invariant pair of b-points.*

Each G'_6 contains transformations of period 3, which fix the b-points individually, and transformations of period 2, which permute the points of the pair.

Several other results along the same lines can be deduced. I will only mention one more.

Every S_2 commutes with exactly four other S_2s, and with exactly four G_3s. This implies that:

Under a transformation of period two, *in addition to the four individually fixed c-points*, *there are also four quadruples of c-points and four pairs of b-points that are invariant as sets.*

Finally, recall that a G''_{24} contains four G_3s. Accordingly, we get four pairs of related b-points, and by what has been said about the G''_{24}, it is clear that *these four pairs of points are permuted in every possible way by the transformations of the* G''_{24}.

In all the statements above it is implicit that there are no more of each type of invariant set or fully permuted set than the ones stated.

3. The Normal Curve of Order Four

As the variable η in our equation of degree 168 we can choose any algebraic function that is single-valued on the Riemann surface just described and takes 168 distinct values on a generic orbit of the group of transformations. We will in any case want to select the simplest function when it comes to actually constructing the equation, and therefore I will first deal with the problem of finding the *normal curve of lowest order* from which the equation between η and J can be derived. This problem is settled, as we will soon see, by means of a series of simple deductions made possible by the fact that *a lot is known about algebraic functions of genus* $p = 3$.[8]

Regarding the normal curve two types of algebraic functions of genus 3 should be distinguished: *hyperelliptic* and *general*. In the hyperelliptic case the normal curve is a [plane] curve C_5 of the fifth order with a triple point, and in the general case it is a [plane] curve of the fourth order [with no multiple points].[9]

I claim, first of all, that *our normal curve cannot be hyperelliptic*. For our curve must, like the equation between J and η from which it is derived, be mapped to itself by 168 one-to-one transformations forming a group whose structure we already know. But a hyperelliptic curve has a one-parameter family of pairs of points that is invariant under one-to-one transformations (for the curve C_5, which has a triple point, this family is given by the intersection of the curve with rays that go through the triple point). Therefore the pencil of rays emanating from the triple point would be mapped to itself in 168 ways.[10] But a pencil of rays is a rational one-dimensional variety; therefore (by a reasoning that I have often used before) there must exist a group of 168 linear transformations behaving in exactly the same way as the group of transformations of the surface. In particular, there should be no transformation of period greater than seven. But it is well-known that such a group cannot exist.

Thus our normal curve has order four.

Now the theory of algebraic functions[11] says that in general, under a one-to-one transformation of a curve to itself, what Riemann called the φ functions transform linearly. For a curve of fourth order, the φ functions take a given value at (generically) four points, and the quadruples thus determined can be regarded as the intersections of the curve with lines going through a certain point of the plane. Thus every linear transformation of φ gives rise to a map of the plane

---

[8] See [Weber 1876].

[9] [See the exposition in [Clebsch and Gordan 1866, p. 65] and in [Clebsch 1876, vol. 1, pp. 687, 712]. –K.]

[10] Conceivably some S_2 might interchange the two intersection points on each ray, so there would be only 84 transformations of the pencil of rays, not 168. The reasoning given in the text would still work, but in any case this possibility does not arise because [such an S_2 would be central and –L.] our group is simple.

[11] See [Brill and Nöther 1874].

that associates to each straight line a straight line and to each point a point—in other words, a *collineation* in the usual sense. Therefore:

Our curve of order four is sent to itself by a group of 168 *collineations, having the structure we already know.* In particular, *there exists a group of* 168 *collineations of the plane none of which has period greater than seven.*[12]

On our curve of order four, most points are grouped in orbits of 168 points each, under the action of these collineations. But there are three smaller orbits: one with only 24 points (which we have labeled a), one with 56 points (labeled b), and one with 84 points (labeled c).

On the other hand, one knows that a curve of order four has distinguished sets with 24, 56 and 84 points: there are 24 *inflection points*, 56 *contact points of bitangents*, and 84 *sextatic points*. [A point is sextatic if some conic makes contact of order six with the curve there. For instance, if a smooth curve has an axis of symmetry — say the y-axis — its intersections with this axis are sextatic points: writing y as a function of x the first, third and fifth derivatives vanish, and among the conics tangent to the surface at the point and symmetric with respect to the same axis we still have two parameters with which to control the second and fourth derivatives. –L.] Each of these sets is characterized by a property that does not change under collineations, and therefore each is invariant, as a set, under our group of 168 collineations. Consequently:

The points a are the inflection points, the points b are the contact points of the bitangents, and the points c are the sextatic points.

One might object that conceivably the inflection points could be a subset of the contact points of bitangents or of the sextatic points, or the last two could be a subset of one another. But this cannot happen because, from what we know about the Riemann surface, only orbits of 24, 56, or 84 points can occur, and 56 is not divisible by 24, nor is 84 a sum of multiples of 24 and 56.

We can also give a simple geometric interpretation to the *triples* of points a, to the *pairs* of points b, and to the *quadruples* of points c.

Regarding the triples, note that every inflection tangent of our curve C_4 intersects the curve in exactly one other point. We thus obtain 24 points, one corresponding to each inflection, and they form an orbit. Since the only 24-point orbit consists of the inflection points themselves, *the intersection points of the inflection tangents coincide with the inflection points, in some permutation.* This permutation cannot fix any inflection point, otherwise the order of contact there would be four, and the inflection points would not all be distinct from one another, nor from the contact points of the bitangents. Thus:

An inflection tangent to the curve C_4 intersects the curve in another inflection point.

[12] This group is missing from the list of all finite groups of linear substitutions in three variables [that is, subgroups of $\mathrm{SL}(3,\mathbb{C})$ –L.] given in [Jordan 1878]. (As Jordan has pointed out to me, the error appears on page 167 of his article, line 8 from below, where Ω need not be divisible by 9φ, only by 3φ. (Added in proof December 1878.))

Now there exist collineations of C_4 into itself that fix a given inflection point a. Such a collineation must also fix the inflection point associated to that a by the process just described, and the inflection point associated to that, and so on. But the *triples* of points a are characterized precisely by the property that a transformation that fixes one point of the triple fixes all three. [See page 294 –L.] We conclude that:

The 24 *inflection points of* C_4 *fall into eight triples*, *corresponding to the triples of* a*-points*. *Each triple of inflection points forms the vertices of a triangle whose edges are the inflection tangents*.

Even simpler is the meaning of the pairs of b-points. If a collineation fixes one of the contact points of a bitangent, it must also fix the other. Therefore:

The 28 *pairs of* b*-points correspond to the* 28 *pairs of contact points of the bitangents*.

Finally, to interpret the quadruples of c-points, we use the easily proved fact that any plane collineation of period two is a perspective transformation. Thus, corresponding to an S_2, we have a perspective *center* and *axis*. There are 21 such centers and axes. Each perspective axis intersects C_4 in four points, and these are the points fixed by the corresponding perspective. Thus:

The 84 *sextatic points are the intersections of* C_4 *with* 21 *straight lines*. *The four points on each of these lines correspond to a quadruple of* c*-points*.

Finally, we revisit the statements given at the end of Section 2. They have the following counterparts:

Each perspective center lies on four perspective axes; *conversely*, *each axis contains four centers*.

Each bitangent contains three centers, *while each center lies on four bitangents*.

The 24 *collineations of a* G''_{24} *permute in all possible ways a certain set of four bitangents*.

4. Equations for the Curve of Order Four

The results already stated are more than enough to allow us to construct, for the curve C_4, several equations for which the different types of collineations stand out.

First, we might choose as our coordinate triangle a *triangle of inflection tangents*. Let its sides be $\lambda = 0$, $\mu = 0$ and $\nu = 0$, with the side $\lambda = 0$ osculating the curve at the intersection with $\mu = 0$, and so on cyclically. Then the equation of the curve must have the form

$$A\lambda^3\mu + B\mu^3\nu + C\nu^3\lambda + \lambda\mu\nu(D\lambda + E\mu + F\nu) = 0.$$

Now the curve is invariant under a cyclic permutation of λ, μ, ν. Replacing λ, μ and ν by appropriate multiples, we can arrange to have $A = B = C$ and $D = E = F$. Next, the curve is sent to itself by six collineations of period seven that leave each side of the triangle invariant. These collineations can be expressed analytically in such a way that the ratios $\lambda : \mu : \nu$ are multiplied by appropriate seventh roots of unity. Such a substitution cannot take the term $\lambda\mu\nu(\lambda+\mu+\nu)$ to a multiple of itself, so this term cannot in fact appear. *Therefore the equation reads simply*

$$0 = f = \lambda^3\mu + \mu^3\nu + \nu^3\lambda. \tag{1}$$

I will always express the collineations that take f to itself in such a way that the determinant is one. We first have the collineation of period three given by

$$\lambda' = \mu, \quad \mu' = \nu, \quad \nu' = \lambda, \tag{2}$$

and then the period-seven collineation

$$\lambda' = \gamma\lambda, \quad \mu' = \gamma^4\mu, \quad \nu' = \gamma^2\nu, \tag{3}$$

where $\gamma = e^{2\pi i/7}$. If we combine these two collineations and their iterates in all possible ways, *we obtain the G'_{21} that leaves invariant the chosen inflection triangle.*

To highlight the six elements of a G'_6 I will choose a new coordinate triangle, whose sides are defined by the property of being each fixed by the permutations (2). Thus I start by setting

$$x_1 = \frac{\lambda+\mu+\nu}{\alpha-\alpha^2}, \quad x_2 = \frac{\lambda+\alpha\mu+\alpha^2\nu}{\alpha-\alpha^2}, \quad x_3 = \frac{\lambda+\alpha^2\mu+\alpha\nu}{\alpha-\alpha^2}, \tag{4}$$

where $\alpha = e^{2\pi i/3}$.

The equation of the curve becomes

$$0 = f = \tfrac{1}{3}\big(x_1^4 + 3x_1^2x_2x_3 - 3x_2^2x_3^2 + x_1((1+3\alpha^2)x_2^3 + (1+3\alpha)x_3^3)\big). \tag{5}$$

To get rid of the cube roots of unity, we further set

$$x_1 = \frac{y_1}{\sqrt[3]{7}}, \quad x_2 = y_2\sqrt[3]{3\alpha+1}, \quad x_3 = y_3\sqrt[3]{3\alpha^2+1} \tag{6}$$

and get

$$0 = f = \frac{1}{21\sqrt[3]{7}}\big(y_1^4 + 21y_1^2y_2y_3 - 147y_2^2y_3^2 + 49y_1(y_2^3+y_3^3)\big). \tag{6a}$$

We immediately see that $y_1 = 0$ is a bitangent, with contacts at $y_2 = 0$ and $y_3 = 0$, and that *the six substitutions of the corresponding G'_6 are generated by*

$$y_1' = y_1, \qquad y_2' = \alpha y_2, \quad y_3' = \alpha^2 y_3, \tag{7}$$

$$y_1' = -y_1, \quad y_2' = -y_3, \quad y_3' = -y_2 \tag{8}$$

(the first of these coincides with (2)).

The three perspective centers lying on $y_1 = 0$ are given by

$$y_2 + y_3 = 0, \quad y_2 + \alpha y_3 = 0, \quad y_2 + \alpha^2 y_3 = 0,$$

while the corresponding perspective axes have the equations

$$y_2 - y_3 = 0, \quad y_2 - \alpha y_3 = 0, \quad y_2 - \alpha^2 y_3 = 0.$$

In order to locate a G''_{24}, I will first of all find the bitangents that go through the perspective centers just listed. Each center lies on four bitangents, but one of them is just the line $y_1 = 0$, so there remain nine bitangents to be found. First we consider those that go through the center $y_1 = y_2 + y_3 = 0$, and which therefore have an equation of the form

$$\sigma y_1 + (y_2 + y_3) = 0.$$

To determine σ, we substitute the value of y_1 from the preceding equation into the equation of the curve, then sort by powers of $y_2 y_3/(y_2 + y_3)^2$ to obtain a quadratic equation in this quantity, and finally set its determinant to zero. We get

$$28\sigma^3 - 21\sigma^2 - 6\sigma - 1 = 0,$$

whose roots are $\sigma = 1$ and $\sigma = \frac{1}{8}(-1 \pm 3\sqrt{-1/7})$. We conclude that *the bitangents that go through the point $y_1 = y_2 + y_3 = 0$ have equations*

$$y_1 + y_2 + y_3 = 0 \quad \textit{and} \quad (-7 \pm 3\sqrt{-7})y_1 + 56y_2 + 56y_3 = 0.$$

The remaining six bitangents (going through the other two centers) are obtained from these three by two applications of (7).

Now I claim that $y_1 = 0$, *together with any three of the bitangents just discussed that are sent to one another by* (7), *form a quadruple of bitangents whose eight contact points lie on a conic.* More generally, the six points in any orbit of the G'_6, together with the contact points of $y_1 = 0$, lie on a conic, because the substitutions (7) and (8) preserve the quadratic expression $y_1^2 + ky_2y_3$, for each k. The preceding claim is a particular case of this fact, because the six contact points form an orbit of the G'_6 (each given bitangent goes through the perspective center of an S_2, and so is invariant under it, its two contact points being interchagned).

In view of this, we can write the equation of our C_4 in three different ways in the form $pqrs - w^2 = 0$, where p, q, r, s are bitangents and w is the conic that goes through the contact points. The first such expression is

$$0 = \frac{1}{21\sqrt[3]{7}}\big(49y_1(y_1+y_2+y_3)(y_1+\alpha y_2+\alpha^2 y_3)(y_1+\alpha^2 y_2+\alpha y_3) - 3(4y_1^2-7y_2y_3)^2\big), \tag{9}$$

and the other two are

$$\begin{aligned}0 = \frac{1}{21\sqrt[3]{7}}\Bigg(\frac{y_1}{7\cdot 8^3}&\big((-7 \pm 3\sqrt{-7})y_1 + 56y_2 + 56y_3\big)\\ &\times \big((-7 \pm 3\sqrt{-7})y_1 + 56\alpha y_2 + 56\alpha^2 y_3\big)\\ &\times \big((-7 \pm 3\sqrt{-7})y_1 + 56\alpha^2 y_2 + 56\alpha y_3\big)\\ &\qquad -3\Big(\frac{1 \pm 3\sqrt{-7}}{16}y_1^2 - 7y_2y_3\Big)^2\Bigg).\end{aligned} \tag{10}$$

Equation (9) will be important in Section 15; the other one *yields, as I will now show, the substitutions in a* G_{24}''. Set

$$\begin{cases} \mathfrak{z}_1 = (21 \mp 9\sqrt{-7})\,y_1, \\ \mathfrak{z}_2 = (-7 \pm 3\sqrt{-7})\,y_1 + \quad 56y_2 + \quad 56y_3, \\ \mathfrak{z}_3 = (-7 \pm 3\sqrt{-7})\,y_1 + \ 56\alpha y_2 + 56\alpha^2 y_3, \\ \mathfrak{z}_4 = (-7 \pm 3\sqrt{-7})\,y_1 + 56\alpha^2 y_2 + \ 56\alpha y_3, \end{cases} \tag{11}$$

so that $\sum \mathfrak{z}_i = 0$. Then (10) becomes, apart from a scalar factor,

$$\left(\sum \mathfrak{z}_i^2\right)^2 - (14 \pm 6\sqrt{-7})\,\mathfrak{z}_1\mathfrak{z}_2\mathfrak{z}_3\mathfrak{z}_4 = 0, \tag{12}$$

and this equation is invariant under the 24 *collineations determined by the permutations of the* $\mathfrak{z}_i$. *These, therefore, are the collineations of the* G_{24}'' *in question.*

We see that the collineations of a G_{24}'' always leave invariant a certain conic

$$\sum \mathfrak{z}_i^2 = 0,$$

which goes through the contact points of the corresponding bitangents. Since there are $2\cdot 7$ groups G_{24}'' and all bitangents have equal title, there are $2\cdot 7$ such conics, and by taking any seven together and intersecting with the curve C_4 we get all the contact points of bitangents. These conics will be very important in the sequel.

5. The 168 Collineations in Relation to the Inflection Triangle. Other Formulas

From (4) and (6) we obtain the following equations connecting the variables λ, μ, ν with y_1, y_2, y_3:

$$\begin{cases} -\sqrt{-3}\sqrt[3]{7}\,\lambda = y_1 + \quad \sqrt[3]{7(3\alpha+1)}\,y_2 + \quad \sqrt[3]{7(3\alpha^2+1)}\,y_3, \\ -\sqrt{-3}\sqrt[3]{7}\,\mu = y_1 + \alpha^2\sqrt[3]{7(3\alpha+1)}\,y_2 + \alpha\ \sqrt[3]{7(3\alpha^2+1)}\,y_3, \\ -\sqrt{-3}\sqrt[3]{7}\,\nu = y_1 + \alpha\ \sqrt[3]{7(3\alpha+1)}\,y_2 + \alpha^2\sqrt[3]{7(3\alpha^2+1)}\,y_3. \end{cases} \tag{12a}$$

If we now apply the substitution (8), replacing y_1, y_2, y_3 by $-y_1, -y_2, -y_3$, we get

$$\begin{aligned} -\sqrt{-3}\sqrt[3]{7}\,\lambda' &= y_1 - \quad \sqrt[3]{7(3\alpha^2+1)}\,y_2 - \quad \sqrt[3]{7(3\alpha+1)}\,y_3, \\ -\sqrt{-3}\sqrt[3]{7}\,\mu' &= y_1 - \alpha\ \sqrt[3]{7(3\alpha^2+1)}\,y_2 - \alpha^2\sqrt[3]{7(3\alpha+1)}\,y_3, \\ -\sqrt{-3}\sqrt[3]{7}\,\nu' &= y_1 - \alpha^2\sqrt[3]{7(3\alpha^2+1)}\,y_2 - \alpha\ \sqrt[3]{7(3\alpha+1)}\,y_3. \end{aligned}$$

Eliminating y_1, y_2, y_3 by combining the two systems, *we obviously find the change from one triangle of inflection tangents*, $\lambda\mu\nu = 0$, *to another*, $\lambda'\mu'\nu' = 0$. The calculation yields a very simple result if we use the well-known expressions for the cube roots on the right-hand sides in terms of third and seventh roots of unity.[13] Setting

$$A = \frac{\gamma^5 - \gamma^2}{\sqrt{-7}}, \quad B = \frac{\gamma^3 - \gamma^4}{\sqrt{-7}}, \quad C = \frac{\gamma^6 - \gamma}{\sqrt{-7}}, \tag{13}$$

$$\sqrt{-7} = \gamma + \gamma^4 + \gamma^2 - \gamma^6 - \gamma^3 - \gamma^5,$$

one easily gets

$$\begin{cases} \lambda' = A\lambda + B\mu + C\nu, \\ \mu' = B\lambda + C\mu + A\nu, \\ \nu' = C\lambda + A\mu + B\nu. \end{cases} \tag{14}$$

If we now combine this substitution (*which has period two*) *in all possible ways with arbitrary iterates of the substitutions* (2) *and* (3),

$$\begin{aligned} &\lambda' = \mu, &&\mu' = \nu, &&\nu' = \lambda, \\ &\lambda' = \gamma\lambda, &&\mu' = \gamma^4\mu, &&\nu' = \gamma^2\nu, \end{aligned}$$

we get in explicit form all the 168 *collineations that preserve our curve of order four*, *or rather*, *the ternary quartic form*

$$f = \lambda^3\mu + \mu^3\nu + \nu^3\lambda.$$

It follows from this result that the coordinates of all the singular elements of our curve can be deduced without further ado: one need only determine the coordinates of one element of the desired kind and apply to them these 168 collineations. In this way it is straightforward to compute the coordinates of the inflection points and corresponding inflection tangents. As for the bitangents, let me remark that the bitangent $y_1 = 0$ of the preceding section, has the equation $\lambda + \mu + \nu = 0$ in terms of our inflection triangle, and that the contact points have coordinates $1 : \alpha : \alpha^2$ and $1 : \alpha^2 : \alpha$. Finally, in order to determine the 21 perspective axes and corresponding centers, it is enough to compute these elements for the substitution (14). We find for the perspective axis

$$\lambda' + \lambda = \mu' + \mu = \nu' + \nu = 0, \tag{15}$$

and for the corresponding perspective center

$$-B - C : B : C \quad \text{or} \quad B : -B - A : A \quad \text{or} \quad C : A : -C - A,$$

all of which indicate the same point.

In the sequel I will mainly use the expressions in λ, μ, ν that, when set to zero, represent the *eight inflection triangles* and the *two times seven conics*, respectively, discussed at the end of the preceding section. I will set down these

[13] $\sqrt[3]{7(3\alpha+1)} = (\gamma+\gamma^6)+\alpha(\gamma^2+\gamma^5)+\alpha^2(\gamma^4+\gamma^3)$ and likewise with α and α^2 interchanged.

equations here as they arise from one another by means of the 168 substitutions of determinant 1.

Denote by δ_∞ the inflection triangle to be used as a coordinate triangle, and write

$$\delta_\infty = -7\lambda\mu\nu, \tag{16}$$

introducing on the right a factor that will later prove convenient. The following formulas then arise for the remaining inflection triangles, where $x = 0, 1, \ldots, 6$:

$$\begin{aligned}\delta_x &= -7(A\gamma^x\lambda + B\gamma^{4x}\mu + C\gamma^{2x}\nu)(B\gamma^x\lambda + C\gamma^{4x}\mu + A\gamma^{2x}\nu)(C\gamma^x\lambda + A\gamma^{4x}\mu + B\gamma^{2x}\nu)\\ &= +\lambda\mu\nu - (\gamma^{3x}\lambda^3 + \gamma^{5x}\mu^3 + \gamma^{6x}\nu^3) + (\gamma^{6x}\lambda^2\mu + \gamma^{3x}\mu^2\nu + \gamma^{5x}\nu^2\lambda)\\ &\qquad + 2(\gamma^{4x}\lambda^2\nu + \gamma^x\nu^2\mu + \gamma^{2x}\mu^2\lambda).\end{aligned} \tag{17}$$

Next we obtain equations for two of the 14 conics, by taking the equation $\sum \mathfrak{z}^2 = 0$ of the preceding section and expressing it first in terms of the y_i and from there in terms of λ, μ, ν:

$$(\lambda^2 + \mu^2 + \nu^2) + \frac{-1 \pm \sqrt{-7}}{2}(\mu\nu + \nu\lambda + \lambda\mu) = 0.$$

Correspondingly, if we denote the left-hand side of the conics by c_x, for $x = 0, 1, 2, \ldots, 6$, we get

$$c_x = (\gamma^{2x}\lambda^2 + \gamma^x\mu^2 + \gamma^{4x}\nu^2) + \frac{-1 \pm \sqrt{-7}}{2}(\gamma^{6x}\mu\nu + \gamma^{3x}\nu\lambda + \gamma^{5x}\lambda\mu). \tag{18}$$

It is these two expressions that will later lead to the simplest resolvents of eighth and seventh degree.

6. Construction of the Equation of Degree 168[14]

As already mentioned, for the role of the variable η in the equation of degree 168 we can choose any single-valued function on our C_4 — and so any rational function of $\lambda : \mu : \nu$ — that takes in general distinct values at the 168 points of an orbit of the group of collineations. It seems simplest to choose λ/μ or λ/ν. But the result gains greatly in clarity if we introduce not *one* such ratio but *both* at once, the two being connected by the equation

$$f = \lambda^3\mu + \mu^3\nu + \nu^3\lambda = 0.$$

For then J can be expressed as a rational function of order 42 of $\lambda : \mu : \nu$,

$$J = R(\lambda, \mu, \nu), \tag{19}$$

where R has a *very simple* form, and the *order*-42 *equation* (19), *together with the order-four equation* $f = 0$, *replaces the* one *degree*-168 *equation that we have*

[14] [Sections 6 through 10 may be compared with [Klein 1879c], which appeared a half year after the present article (March 1879) and is closely connected with it, but unfortunately had to be separated from it in these collected works. –K.]

been talking about so far. A similar procedure appears always to be appropriate when one is dealing with the construction of an equation of genus p greater than zero.

The function $R(\lambda, \mu, \nu)$ must first of all have the property of invariance under the 168 collineations. Thus, to find R, I will first discuss the construction of *all* functions of λ, μ, ν that have this property. Here we assume, of course, that the 168 collineations have been chosen to have determinant *one*. We know *one* such function,

$$f = \lambda^3\mu + \mu^3\nu + \nu^3\lambda;$$

we also know that the *covariants* of f always have the same invariance property. A short argument then shows that the covariants of f can be covered by the desired functions, and allows us at the same time to construct the whole system of functions with the relations that hold between the forms of the system. The rational function R proves to be the simplest combination of dimension zero that can be formed from the covariants. This is the same method that Gordan and I have repeatedly used in our recent works.

The first covariant of f is the Hessian ∇ of order six:

$$\nabla = \frac{1}{54}\begin{vmatrix} \dfrac{\partial^2 f}{\partial\lambda^2} & \dfrac{\partial^2 f}{\partial\lambda\,\partial\mu} & \dfrac{\partial^2 f}{\partial\lambda\,\partial\nu} \\ \dfrac{\partial^2 f}{\partial\mu\,\partial\lambda} & \dfrac{\partial^2 f}{\partial\mu^2} & \dfrac{\partial^2 f}{\partial\mu\,\partial\nu} \\ \dfrac{\partial^2 f}{\partial\nu\,\partial\lambda} & \dfrac{\partial^2 f}{\partial\nu\,\partial\mu} & \dfrac{\partial^2 f}{\partial\nu^2} \end{vmatrix} = 5\lambda^2\mu^2\nu^2 - (\lambda^5\nu + \nu^5\mu + \mu^5\lambda). \tag{20}$$

When set equal to zero, this expression determines the 24 inflection points on the surface $f = 0$, which indeed form an orbit. Now, there was no other orbit having 24 points on $f = 0$, and none having fewer. We conclude that there can be no invariant polynomial function of order less than six, and that any invariant polynomial function of order six must be a multiple of ∇. For if there were another function of order six, it would be expressible in the form

$$k\nabla + l\varphi f,$$

where k and l are constants — since when set to zero it must determine on $f = 0$ the same 24 inflection points. Here φ would be an invariant function of degree two, and such a function, as already remarked, does not exist. In exactly the same way we conclude that *the next higher invariant polynomial function has degree* 14 *and*, *when set to zero*, *it determines on* $f = 0$ *the* 56 *contact points of the bitangents*.

There are different ways in which a covariant of order 14 can be constructed. As is well known, Hesse has constructed for a general curve of order four a curve of order 14 that goes through the contact points of the bitangents. In our case this property holds for *any* covariant of order 14 that is not a multiple of $f^2\nabla$, so we can choose any of them. I choose

$$C = \frac{1}{9}\begin{vmatrix} \dfrac{\partial^2 f}{\partial\lambda^2} & \dfrac{\partial^2 f}{\partial\lambda\,\partial\mu} & \dfrac{\partial^2 f}{\partial\lambda\,\partial\nu} & \dfrac{\partial\nabla}{\partial\lambda} \\ \dfrac{\partial^2 f}{\partial\mu\,\partial\lambda} & \dfrac{\partial^2 f}{\partial\mu^2} & \dfrac{\partial^2 f}{\partial\mu\,\partial\nu} & \dfrac{\partial\nabla}{\partial\mu} \\ \dfrac{\partial^2 f}{\partial\nu\,\partial\lambda} & \dfrac{\partial^2 f}{\partial\nu\,\partial\mu} & \dfrac{\partial^2 f}{\partial\nu^2} & \dfrac{\partial\nabla}{\partial\nu} \\ \dfrac{\partial\nabla}{\partial\lambda} & \dfrac{\partial\nabla}{\partial\mu} & \dfrac{\partial\nabla}{\partial\nu} & 0 \end{vmatrix} = (\lambda^{14} + \mu^{14} + \nu^{14}) + \cdots. \qquad (21)$$

I also form *a function of degree* 21, *the functional determinant of* f, ∇, *and* C:

$$K = \frac{1}{14}\begin{vmatrix} \dfrac{\partial f}{\partial\lambda} & \dfrac{\partial\nabla}{\partial\lambda} & \dfrac{\partial C}{\partial\lambda} \\ \dfrac{\partial f}{\partial\mu} & \dfrac{\partial\nabla}{\partial\mu} & \dfrac{\partial C}{\partial\mu} \\ \dfrac{\partial f}{\partial\nu} & \dfrac{\partial\nabla}{\partial\nu} & \dfrac{\partial C}{\partial\nu} \end{vmatrix} = -(\lambda^{21} + \mu^{21} + \nu^{21}) + \cdots. \qquad (22)$$

When set equal to zero, K determines the 84 sextatic points on $f = 0$. Again, one can infer that apart from K there is no invariant function of order 21; for if there were it would be expressible in the form

$$k K - l \varphi f^\nu,$$

where k and l are constants. Here φ would be an invariant function of degree $21 - 4\nu$, and so when set to zero it would intersect $f = 0$ in a number of points divisible by 4 but not by 8. But the only eligible orbits have 24 or 56 elements; this yields a contradiction.

Now recall that earlier we found the 84 sextatic points as the intersections of $f = 0$ with 21 straight lines, *the* 21 *perspective axes*; see (15). Therefore:

The equation $K = 0$ *represents the union of the* 21 *axes.*

If one wants to determine on $f = 0$ a *general* orbit of 168 points, it is clearly sufficient to consider the pencil of curves

$$\nabla^7 = k\, C^3,$$

for varying k. From this it follows first of all that *under the condition* $f = 0$ *we have*, *for appropriate values of* k *and* l, *a relation of the form*

$$\nabla^7 = k\, C^3 + l K^2; \qquad (23)$$

and then it follows further that f, ∇, C, *and* K, *which are connected by this one equation*, *generate the whole system of forms under consideration*, *and a fortiori the whole system of covariants of* f.

To determine the constants k and l that appear in (23), I start by setting $\lambda = 1$, $\mu = 0$, $\nu = 0$. Formulas (20), (21), (22) yield

$$\nabla = 0, \quad C = 1, \quad K = -1, \tag{23a}$$

and moreover $f = 0$. Thus

$$k = -l.$$

Next I take f in the form (6a),

$$f = \frac{1}{21\sqrt[3]{7}}\big(y_1^4 + 21y_1^2y_2y_3 - 147y_2^2y_3^2 + 49y_1(y_2^3 + y_3^3)\big),$$

and compute some terms of ∇, C, and K, obtaining

$$\nabla = \tfrac{1}{27}(7^2y_3^6 - 3\cdot5\cdot7y_1y_2y_3^4\ldots), \quad C = \frac{2^3\cdot7^5\cdot\sqrt[3]{7}}{3^6}y_2y_3^{13}\ldots, \quad K = \frac{-2^3\cdot7^7}{3^9}y_3^{21}\ldots.$$

Now put $y_1 = 0$, $y_2 = 0$, $y_3 = 1$ in these equations, to obtain, besides $f = 0$,

$$\nabla = \frac{7^2}{3^3}, \quad C = 0, \quad K = -\frac{2^3\cdot7^7}{3^9}, \tag{23b}$$

so that

$$l = \frac{1}{2^6\cdot3^3}, \quad k = \frac{-1}{2^6\cdot3^3}.$$

Thus the relation among ∇, C *and* K *is*

$$(-\nabla)^7 = \left(\frac{C}{12}\right)^3 - 27\left(\frac{K}{216}\right)^2. \tag{24}$$

Based on this relation the function $R(\lambda, \mu, \nu) = J$ can now be determined immediately. J must be equal to 0 at the contact points of the bitangents, equal to 1 at the sextatic points, and equal to ∞ at the inflection points; it should take any other value on some 168-point orbit, and only there. *Thus we have the equation*

$$J : J{-}1 : 1 = \left(\frac{C}{12}\right)^3 : 27\left(\frac{K}{216}\right)^2 : -\nabla^7, \tag{25}$$

and this equation, together with $f = 0$, represents the problem of degree 168 that we had set out to formulate.

If we use, instead of J, the invariants g_2, g_3, Δ of elliptic integrals, we can write

$$g_2 = \frac{C}{12}, \quad g_3 = \frac{K}{216}, \quad \sqrt[7]{\Delta} = -\nabla. \tag{26}$$

7. Lower-Degree Resolvents

The group of 168 collineations contains subgroups G'_{21} and G'_{24} of 21 and 24 elements. Accordingly, our problem of degree 168 has resolvents of degree eight and of degree seven. There can be no question about what is the simplest form of these resolvents; they must be exactly the equations of degree eight and seven given respectively in [Klein 1879a, Equation (15)] and [Klein 1879b, Equations (5) to (7)], and which I constructed directly starting from the ω-substitutions. So all that is left to find out is how to pass from our current description to the equations given in those earlier articles. As always in such cases, this can be done in two ways.

The first approach is to seek the simplest *rational* function $r(\lambda, \mu, \nu)$ that takes the same value at all the points of any orbit of the subgroup G'_{21} or G'_{24} under consideration, and then ask how this function is related to J.

The second is to find the lowest-degree *polynomial* function of λ, μ, ν that remains invariant under the substitutions in the desired subgroup, and then determine its relationship with ∇, C, K or with Δ, g_2, g_3.

Each method has its advantages, and in the sequel we use the second to complement the first.

8. The Resolvent of Degree Eight

Consider the G'_{21} generated by the two substitutions

$$\lambda' = \mu, \qquad \mu' = \nu, \qquad \nu' = \lambda,$$
$$\lambda' = \gamma\lambda, \qquad \mu' = \gamma^4\mu, \qquad \nu' = \gamma^2\nu.$$

It leaves invariant the inflection triangle $\delta_\infty = -7\lambda\mu\nu$ of (16), and of course ∇, so also the rational function $\sigma = \delta_\infty^2/\nabla$. The latter has the property that it takes a prescribed value at only 21 points of the curve $f = 0$, because the pencil of order-six curves $\delta_\infty^2 - \sigma\nabla$ has three fixed points (the vertices of the coordinate triangle) in common with $f = 0$, each with multiplicity one. Thus, if we use σ as a variable, J *becomes a rational function of* σ, *of degree eight*:

$$J = \frac{\varphi(\sigma)}{\psi(\sigma)}. \tag{27}$$

Now we determine the multiplicity of the individual factors in φ, ψ, and $\varphi - \psi$, as I have done several times in similar problems.

J becomes infinite with multiplicity seven at the 24 inflection points. At three of these points — the vertices of the coordinate triangle — σ vanishes with multiplicity seven, since δ_∞ has a fourfold zero and ∇ a simple zero. At the remaining 21 inflection points σ becomes infinite with multiplicity one because of the denominator ∇. Therefore $\psi(\sigma)$ consists of a simple factor and a sevenfold one, the first vanishing at $\sigma = 0$ and the second at $\sigma = \infty$. *Thus, apart from a constant factor*, $\psi(\sigma)$ *equals* σ.

J vanishes with multiplicity three at the 56 contact points of the bitangents. With respect to the group G'_{21} the bitangents fall into two classes, one with 7 and one with 21 elements,[15] so the contact points are divided into two orbits with 7 points each and 2 with 21 each. At points of the first kind σ takes a certain value with multiplicity three, and at points of the second with multiplicity one. This means that φ *contains two simple and two threefold linear factors*.

Finally, J takes the value 1 with multiplicity two at the 84 sextatic points. With respect to G'_{21} these points fall into four orbits of 21, and at each point σ takes its value with multiplicity one. Therefore $\varphi - \psi$ *is the square of an expression of degree four of nonzero discriminant*.

Now these are the same conditions on φ, ψ, $\varphi - \psi$ that led me in [Klein 1879a, Abschnitt II, § 14] to the construction of the modular equation of degree eight:

$$J : J-1 : 1 = (\tau^2+13\tau+49)(\tau^2+5\tau+1)^3 : (\tau^4+14\tau^3+63\tau^2+70\tau-7)^2 : 1728\tau. \tag{28}$$

We arrive at this same equation in the present case, if we denote an appropriate multiple of σ *by* τ.

To determine this multiple, I now return to the y-coordinate system of (12a). The value of $7^2\lambda^2\mu^2\nu^2$ is $(5-3\alpha)\cdot 7^2/3^3$ when $y_1 = 0$, $y_2 = 0$, $y_3 = 1$, and $(5-3\alpha^2)\cdot 7^2/3^3$ when $y_1 = 0$, $y_2 = 1$, $y_3 = 0$. In both cases $\nabla = 7^2/3^3$ by (23b), so σ has the values $(5-3\alpha)$ and $(5-3\alpha^2)$, respectively. But the points $(y_1, y_2, y_3) = (0,0,1)$ and $(0,1,0)$ are the contact points of the bitangent $\lambda + \mu + \nu = 0$, which is one of the seven distinguished bitangents with respect to the chosen G'_{21}. Accordingly, J vanishes at these points and in particular the simple factor $\tau^2 + 13\tau + 49$ in (28) also vanishes. Its roots equal $3\alpha - 5$ and $3\alpha^2 - 5$. *Therefore we have simply*

$$\tau = -\sigma,$$

or, put another way:

One root τ *of Equation* (28) *has the value*

$$\tau_\infty = -\frac{\delta_\infty^2}{\nabla} = -\frac{7^2\lambda^2\mu^2\nu^2}{\nabla}. \tag{29}$$

Then (17) implies that *the remaining roots* τ_x *have the values*

$$\tau_x = -\frac{\delta_x^2}{\nabla} = -\frac{\left(\begin{aligned}\lambda\mu\nu - (\gamma^{3x}\lambda^3 + \gamma^{5x}\mu^3 + \gamma^{6x}\nu^3) + (\gamma^{6x}\lambda^2\mu + \gamma^{3x}\mu^2\nu + \gamma^{5x}\nu^2\lambda)\\ + 2(\gamma^{4x}\lambda^2\nu + \gamma^{x}\nu^2\mu + \gamma^{2x}\mu^2\lambda)\end{aligned}\right)^2}{\nabla}. \tag{30}$$

[15] This and similar statements can be verified easily using the formulas given earlier.

and so, as promised in the introduction, we have expressed the roots of the modular equation of degree eight as a rational function of one *point on the curve* $f = 0$.

As mentioned in [Klein 1879a, Abschnitt II, § 18], Equation (28) can be transformed as follows: write z^2 instead of τ, $27g_3^2/\Delta$ instead of $J-1$, and take the square root of both sides, to obtain

$$z^8 + 14z^6 + 63z^4 + 70z^2 - \frac{216g_3}{\sqrt{\Delta}}z - 7 = 0. \tag{31}$$

We can further replace $216g_3/\sqrt{\Delta}$ with $K/\sqrt{-\nabla^7}$, by (26); and replacing also z by its value $\delta/\sqrt{-\nabla}$, given by (29), (30), the result is

$$\delta^8 - 14\delta^6\nabla + 63\delta^4\nabla^2 - 70\delta^2\nabla^3 - \delta K - 7\nabla^4 = 0. \tag{32}$$

To see that the penultimate term should appear with a negative sign one can, for example, set $(\lambda, \mu, \nu) = (1, 0, 0)$ and replace δ by any of the values δ_x.

We would have arrived at the same equation (32) *if we had taken the polynomial approach.* For the simplest *polynomial* function of λ, μ, ν that is left invariant by G'_{21} is $\delta_\infty = -7\lambda\mu\nu$. Under the 168 collineations δ takes eight distinct values, whose symmetric function must be a polynomial function of ∇, C, K (since f is taken to equal 0). Therefore δ satisfies an equation of the eighth degree, which, in view of the degrees of ∇, C, K, must have the form

$$\delta^8 + a\nabla\delta^6 + b\nabla^2\delta^4 + c\nabla^3\delta^2 + dK\delta + e\nabla^4 = 0,$$

and if the coefficients a, b, c, d, e are determined by substituting for δ, ∇, K their values in terms of λ, μ, ν and taking into account that $f = 0$, we recover (32). This derivation has the advantage that it shows a priori why only certain powers of δ appear in (32).

9. Contact Curves of the Third Order. Solution of the Equation of Degree 168.

The eight roots of (32) can be expressed as follows, by virtue of (16) and (17):

$$\begin{cases} \delta_\infty = -7\lambda\mu\nu, \\ \delta_x \;= \lambda\mu\nu - \gamma^{-x}(\nu^3 - \lambda^2\mu) - \gamma^{-4x}(\lambda^3 - \mu^2\nu) - \gamma^{-2x}(\mu^3 - \nu^2\lambda) \\ \qquad\qquad + 2\gamma^{x}\nu^2\mu \qquad\qquad + 2\gamma^{4x}\lambda^2\nu \qquad\qquad + 2\gamma^{2x}\mu^2\lambda. \end{cases} \tag{33}$$

Now, I have already stated in [Klein 1879a, Abschnitt II, end of § 18] that Equation (31), and therefore also (32), is a *Jacobian equation of degree eight*, that is, the square roots of its roots can be written in terms of four quantities

A_0, A_1, A_2, A_3 as follows:[16]

$$\begin{cases} \sqrt{\delta_\infty} = \sqrt{-7}\,A_0, \\ \sqrt{\delta_x} \;= \qquad A_0 + \gamma^{\rho x} A_1 + \gamma^{4\rho x} A_2 + \gamma^{2\rho x} A_3 \end{cases} \tag{34}$$

(where ρ is any integer not divisible by 7). One may ask how this assertion, which I had deduced from the transcendent solution of (31) (loc. cit.) can be verified algebraically. This is done by considering certain *contact curves or order three*[17] of our curve $f = 0$, or, in other words, by considering *certain root functions of order three that exist on the curve* $f = 0$.

It is known that a curve of order four possesses 64 triply infinite families of contact curves of order three, of which 36 have *even* and 28 *odd* characteristic.[18] *In our case one family of even characteristic is singled out by the fact that it contains the eight inflection triangles as contact curves.*

We can certainly regard an inflection triangle as a contact curve of third order, in that its 12 intersections with our curve of order four actually coalesce into only *three* points, four at a time. Now consider, say, the triangle δ_∞. Through its intersections with C_4 we place the triply infinite family of curves of third order that have contact with C_4 at those points; their equation is

$$k\,\lambda\mu\nu + a\lambda^2\mu + b\mu^2\nu + c\nu^2\lambda = 0. \tag{35}$$

Each cubic meets the C_4 in another six points, and it is well-known that these are the contact points of another contact cubic belonging to the same family as δ_∞; and in this way one obtains *all* the cubics in the family. Now we have the identity

$$\begin{aligned}(k\,\lambda\mu\nu + a\lambda^2\mu + b\mu^2\nu + c\nu^2\lambda)^2 - (a^2\lambda\mu + b^2\mu\nu + c^2\nu\lambda)f \\ = \lambda\mu\nu\big(k^2\lambda\mu\nu - (a^2\mu^3 + b^2\nu^3 + c^2\lambda^3) + 2(bc\mu\nu^2 + ca\nu\lambda^2 + ab\lambda\mu^2) \\ + ((2ak - b^2)\lambda^2\mu + (2bk - c^2)\mu^2\nu + (2ck - a^2)\nu^2\lambda)\big).\end{aligned} \tag{36}$$

Therefore the totality of the contact cubics in this family is represented by the equation

$$\begin{aligned}0 = k^2\lambda\mu\nu - (a^2\mu^3 + b^2\nu^3 + c^2\lambda^3) + 2(bc\mu\nu^2 + ca\nu\lambda^2 + ab\lambda\mu^2) \\ + \big((2ak - b^2)\lambda^2\mu + (2bk - c^2)\mu^2\nu + (2ck - a^2)\nu^2\lambda\big).\end{aligned} \tag{37}$$

[16] Regarding the degree-eight Jacobian equation, see [Brioschi 1868] and the commentary in [Jung and Armenante 1869], as well as a remark at the end of [Klein 1878a] not yet used in the text [and also [Brioschi 1879]]. I hope to return to subject in detail soon. [See [Klein 1879c].] [See also the recent [Brioschi 1878/79].]

[17] That is, curves of order three that have six first-order contacts with $f = 0$. [The developments in the text follow the investigations in [Hesse 1855].]

[18] [See [Riemann 1861/62] and Section 15 in this article.]

If we set

$$k = 1, \quad a = \gamma^{-x}, \quad b = \gamma^{-4x}, \quad c = \gamma^{-2x},$$

the right-hand side becomes the expression of δ_x, showing that all eight inflection triangles belong to the same family of contact cubics, as claimed.[19]

Now the formulas in (34) easily follow from the statement that *the set of root functions for a family of even characteristic is linearly generated by four independent elements.* Indeed, choose the four root functions correponding to the curves (37) for which, in turn,

$$\begin{aligned} &k = 1, \quad a = 0, \quad b = 0, \quad c = 0,\\ &k = 0, \quad a = 1, \quad b = 0, \quad c = 0, \quad \text{etc.,} \end{aligned}$$

and accordingly set

$$\begin{cases} A_0 = \sqrt{\lambda\mu\nu}, \\ A_1 = \sqrt{-\mu^3 - \nu^2\lambda}, \quad A_2 = \sqrt{-\nu^3 - \lambda^2\mu}, \quad A_3 = \sqrt{-\lambda^3 - \mu^2\nu}. \end{cases} \tag{38}$$

Then, by choosing the signs appropriately and using the condition $f = 0$, one obtains

$$\begin{cases} A_0A_1 = \lambda^2\mu, \quad A_0A_2 = \mu^2\nu, \quad A_0A_3 = \nu^2\lambda, \\ A_1A_2 = \lambda\mu^2, \quad A_2A_3 = \mu\nu^2, \quad A_2A_1 = \nu\lambda^2, \end{cases} \tag{39}$$ [20]

and Equation (37) can be written in the following irrational form:

$$kA_0 + aA_1 + bA_2 + cA_3 = 0. \tag{40}$$

In particular, taking (33) into account,

$$\begin{cases} \sqrt{\delta_\infty} = \sqrt{-7}\, A_0, \\ \sqrt{\delta_x} = \qquad A_0 + \gamma^{-x}A_1 + \gamma^{-4x}A_2 + \gamma^{-2x}A_3. \end{cases} \tag{41}$$

These are exactly the same formulas as (34), *except that the formerly unspecified integer ρ now has been set to* -1.

---

[19] That the family has *even* characteristic follows from the irrational form of its equation, which we are about to state.

[20] Consequently A_0, A_1, A_2, A_3 satisfy a series of identities, all of which can be obtained by setting to zero the determinants of the 3×3 minors of

$$\begin{pmatrix} A_1 & A_0 & -A_2 & 0 \\ A_2 & 0 & A_0 & -A_3 \\ A_3 & -A_1 & 0 & A_0 \end{pmatrix}.$$

One can use these formulas to solve our equation of degree 168 explicitly in terms of elliptic functions.[21] The roots δ of (32) are proportional to the roots z of (31), and for the latter an expression in $q = e^{i\pi\omega}$ was given in [Klein 1879a, Abschnitt II, §§ 17, 18]. Using that expression we obtain here

$$\delta_\infty : \delta_x = -7\sqrt[6]{q^7}\prod(1-q^{14n})^2 : \sqrt[6]{\gamma^x q^{1/7}}\prod(1-\gamma^{2nx}q^{2n/7})^2. \tag{42}$$

The products on the right can be rewritten using the series development

$$q^{1/12}\prod(1-q^{2n}) = \sum_{-\infty}^{\infty}(-1)^n q^{(6n+1)^2/12},$$

so we can write the ratios among A_0, A_1, A_2, A_3 in terms of these series; using the equations

$$\frac{\lambda}{\mu} = \frac{A_0}{A_2}, \quad \frac{\mu}{\nu} = \frac{A_0}{A_3}, \quad \frac{\nu}{\lambda} = \frac{A_0}{A_1}, \tag{43}$$

arising from (39), *we obtain the following solutions for the equation of degree* 168:

$$\begin{cases} \dfrac{\lambda}{\mu} = q^{4/7}\dfrac{\sum_{-\infty}^{\infty}(-1)^{h+1}q^{21h^2+7h}}{\sum_{-\infty}^{\infty}(-1)^h q^{21h^2+h} + \sum_{-\infty}^{\infty}(-1)^h q^{21h^2+13h+2}}, \\ \dfrac{\mu}{\nu} = q^{2/7}\dfrac{\sum_{-\infty}^{\infty}(-1)^{h+1}q^{21h^2+7h}}{\sum_{-\infty}^{\infty}(-1)^{h+1} q^{21h^2+19h+4} + \sum_{-\infty}^{\infty}(-1)^h q^{21h^2+37h+16}}, \\ \dfrac{\nu}{\lambda} = q^{1/7}\dfrac{\sum_{-\infty}^{\infty}(-1)^{h+1}q^{21h^2+7h}}{\sum_{-\infty}^{\infty}(-1)^h q^{21h^2+25h+7} + \sum_{-\infty}^{\infty}(-1)^{h+1} q^{21h^2+31h+11}}. \end{cases} \tag{44}$$

It suffices to compute this one solution, since the other 167 *can be obtained from this one by applying the collineations of Section* 5.

Here I have only computed the ratios $\lambda : \mu : \nu$; if one wishes to start from the formulation represented by Equation (26), one of course gets formulas for the *actual values* of λ, μ, ν.

[21] The equation should also be solvable by means of a linear differential equation of third order; how is the latter to be constructed? [In the *Abhandlungen* this is complemented by a reference to a long footnote to [Klein 1879c], which reads in part as follows: The corresponding differential equation for $f = 0$ has been constructed by Halphen in a letter that reached me on 11 June 1884 [Halphen 1884] and later by Hurwitz [1886]. Let J be as in the text and set $\eta_i = y_i\nabla^8/(C^2K)$, for $i = 1, 2, 3$; then, according to Hurwitz, the η_i are certain solutions of

$$J^2(J-1)^2\frac{d^3\eta}{dJ^3} + (7J-4)J(J-1)\frac{d^2\eta}{dJ^2} + \Big(\tfrac{72}{7}(J^2-J) - \tfrac{20}{9}(J-1) + \tfrac{3}{4}J\Big)\frac{d\eta}{dJ} + \Big(\tfrac{792}{7^3}(J-1) + \tfrac{5}{8} + \tfrac{2}{63}\Big) = 0.$$

–L.]

10. The Resolvent of Degree Seven

The substitutions of a G''_{24} always leave invariant a conic c_x that goes through the contact points of four bitangents. By (18) we can write

$$c_x = (\gamma^{2x}\lambda^2 + \gamma^x\mu^2 + \gamma^{4x}\nu^2) + \frac{-1 \mp \sqrt{-7}}{2}(\gamma^{6x}\mu\nu + \gamma^{3x}\nu\lambda + \gamma^{5x}\lambda\mu). \tag{45}$$

Now form the rational function

$$\xi = \frac{c_x^3}{\nabla}. \tag{46}$$

Since the numerator and denominator are invariant under the substitutions in the G''_{24}, and since the pencil of sixth-order curves $\nabla - \xi\, c_x^3 = 0$ has no fixed intersection with the curve C_4, we conclude that ξ takes a given value at exactly the points of an orbit of the G''_{24}. Therefore:

J is a rational function of degree seven of ξ:

$$J = \frac{\varphi(\xi)}{\psi(\xi)}. \tag{47}$$

We now consider again the values $J = \infty, 0, 1$.

The 24 inflection points, where J becomes infinite with multiplicity seven, form a single orbit of the G''_{24}, each point appearing once. Thus $\psi(\xi)$ is the seventh power of a linear factor. But ξ is itself infinite at the inflection points, because of (46). *Therefore $\psi(\xi)$ is a constant.*

Of the 56 contact points of the 28 bitangents eight lie on $c_x = 0$, so ξ vanishes with order three at those points. The other 48 split into 2 orbits of as 24 (each corresponding to 12 tangents). Thus φ *contains the simple factor ξ and the cube of a quadratic factor of nonzero discriminant.*

Finally, the 84 sextatic points fall into three orbits of 12 points each and two of 24 points each. Thus $\varphi - \psi$ *contains a simple cubic factor and the square of a quadratic factor.*

Again, these are the requirements on φ and ψ that led in [Klein 1879b, § 7] to the construction of the simplest equation of degree seven, which has the following form:

$$\begin{aligned}
J : J-1 : 1 &= \mathfrak{z}\big(\mathfrak{z}^2 - 2^2{\cdot}7^2(7 \mp \sqrt{-7})\mathfrak{z} + 2^5{\cdot}7^4(5 \mp \sqrt{-7})\big)^3 \\
&\quad : \big(\mathfrak{z}^3 - 2^2{\cdot}7{\cdot}13(7 \mp \sqrt{-7})\mathfrak{z}^2 + 2^6{\cdot}7^3(88 \mp 23\sqrt{-7})\mathfrak{z} - 2^8{\cdot}3^3{\cdot}7^4(35 \mp 9\sqrt{-7})\big) \\
&\qquad \times \big(\mathfrak{z}^2 - 2^4{\cdot}7(7 \mp \sqrt{-7})\mathfrak{z} + 2^5{\cdot}7^3(5 \mp \sqrt{-7})\big)^2 \\
&\quad : \mp 2^{27}{\cdot}3^3{\cdot}7^{10}\sqrt{-7}.
\end{aligned} \tag{48}$$

We conclude that the variable $\mathfrak{z}$ coincides with ξ apart from a multiplicative factor, though it is still in question whether the upper sign in front of the $\sqrt{-7}$ in (45) *corresponds to the upper or the lower sign in* (48).

To eliminate this ambiguity, I first transform (48) by setting $\mathfrak{z} = z^3$ and $J = g_2^3/\Delta$, and then I take the cube root of both sides:

$$z^7 - 2^2\cdot 7^2(7 \mp \sqrt{-7})z^4 + 2^5\cdot 7^4(5 \mp \sqrt{-7}) \mp 2^9\cdot 3\cdot 7^3\sqrt{-7}\,\frac{g_2}{\sqrt[3]{\Delta}} = 0. \tag{49}$$

Now, following (26), we substitute $C/(12\sqrt[3]{-\nabla^7})$ for $g_2/\sqrt[3]{\Delta}$ and $kc/\sqrt[3]{\nabla}$ for z, where k is a constant to be determined; the result is

$$k^7c^7 - 2^2\cdot 7^2(7 \mp \sqrt{-7})k^4\nabla c^4 + 2^5\cdot 7^4(5 \mp \sqrt{-7})k\nabla^2 c \pm 2^7\cdot 7^3\sqrt{-7}\,C = 0. \tag{50}$$

This gives

$$\begin{aligned} k^3 \textstyle\sum c_x^3 &= 3\cdot 2^2\cdot 7^2(7 \mp \sqrt{-7})\big(5\lambda^2\mu^2\nu^2 - (\lambda^5\nu + \nu^5\mu + \mu^5\lambda)\big), \\ k^7 \textstyle\prod c_x &= \mp 2^7\cdot 7^3\sqrt{-7}\,(\lambda^{14} + \mu^{14} + \nu^{14} + \cdots) \end{aligned}$$

(naturally under the assumption that $f = 0$). *So the two equations are reconciled when we choose*

$$k = \pm 2\sqrt{-7}$$

and make the sign of k correspond to the upper sign in (49) *and to the lower sign in* (45).

In other words: *The roots z of* (49) *and $\mathfrak{z}$ of* (48) *have the following values in terms of λ, μ, ν:*

$$z = \mathfrak{z}^{1/3} = \frac{\pm 2\sqrt{-7}\Big((\gamma^{2x}\lambda^2 + \gamma^x\mu^2 + \gamma^{4x}\nu^2) + \dfrac{-1\mp\sqrt{-7}}{2}(\gamma^{6x}\mu\nu + \gamma^{3x}\nu\lambda + \gamma^{5x}\lambda\mu)\Big)}{\sqrt[3]{\nabla}}, \tag{51}$$

and so, as promised in the introduction, we have explicitly written the $\mathfrak{z}$'s as rational functions of one point on our C_4.

Equation (50) becomes

$$c^7 + \tfrac{7}{2}(-1 \mp \sqrt{-7})\nabla c^4 - 7\Big(\frac{5 \mp \sqrt{-7}}{2}\Big)\nabla^2 c - C = 0.\text{[22]} \tag{52}$$

Naturally, the polynomial approach would have led to the same equation. Indeed, the lowest polynomial function of λ, μ, ν that remains invariant under a G_{24}'' is exactly the corresponding c_x, and this c_x must satisfy an equation of degree seven, whose coefficients are polynomials in ∇, C, K, and which therefore has the form

$$c^7 + \alpha\nabla c^4 + \beta\nabla^2 c + \gamma C = 0,$$

where α, β, γ are to be determined by the substitution of values for λ, μ, ν. Again, this approach has the advantage of showing a priory a great number of terms must be absent from (52) and (49).

[22] [The corresponding equation for $f \neq 0$ is given in [Klein 1879c, (12)].]

11. Replacement of the Riemann Surface of Section 2 by a Regularly Tiled Cover

Now I would like to explain the relationship between the irrationality $\lambda:\mu:\nu$ and the absolute invariant J, as well as with the roots τ and $\mathfrak{z}$ of the eighth- and seventh-degree equations, in as visual and intuitive a way as possible, using topology.

First recall the figures appearing in [Klein 1879a] for the eight-degree equation and in [Klein 1879b] for the seventh-degree equation. [They are reproduced on the next two pages. –L.] I will start with a general explanation concerning *Riemann surfaces that are related to their Galois resolvent by a rational parameter* [Klein 1879a, Abschnitt III]. Let $F(\eta,z)=0$ be such a surface of degree N; by definition, it has the property that each root η_i is ramified with respect to the parameter z exactly like any other root η_k, so the surface is mapped to itself by N one-to-one transformations (compare Section 2).[23]

We regard the complex values of z as laid out on the plane, and denote by z_1, z_2, ..., z_n the branch points. The branching is the same for all sheets; assume that the sheets come together ν_1 at a time at z_1, ν_2 at a time at z_2, and so on. Now draw on the z-plane any simple closed curve that goes once through each of z_1, z_2, ..., z_n — in other words, a branch cut. It divides the z-plane and each of the N sheets of the η Riemann surface stretched over it into two regions. We think of the first region as being shaded, the second unshaded. Then transform the surface, which lay in sheets above the z-plane, so it now sits in space and is smoothly curved; but maintain the shading and the connectivity of the regions. The resulting surface is therefore divided into $2N$ alternately shaded and unshaded n-gons, which meet at the various vertices in groups of $2\nu_1$, $2\nu_2$, ..., $2\nu_n$ wedges, and which are, in the topological sense, alternately identical with and the mirror image of a given polygon; the edges of the polygons are the images of the branch cut we drew on the z-plane. The N one-to-one transformations of the equation $F(\eta,z)=0$ into itself are reflected in that the surface thus obtained can be mapped one-to-one onto itself in N ways. Indeed, fix a (say) shaded polygon of the surface and map it to any chosen shaded polygon [preserving the numbering of the vertices –L.]; then declare that adjacent polygons should map to adjacent polygons. This assigns to each polygon a unique image in a consistent way, and the resulting correspondence of polygons is determined by the initial choice of an image for the base polygon. I will call covers that are divided in this sense into alternating regions *regularly* [*symmetric*] *tiled covers;* they comprise as particular cases, when the genus p is zero, the tilings of the sphere into 24, 48, and 120 triangles, associated with the tetrahedron, octahedron, and icosahedron.

We can state the following general theorem:

Any Galois resolvent $F(\eta,z)=0$ *admits a regularly tiled cover*.

[23] [See footnote 7 on page 293. –B.-H.]

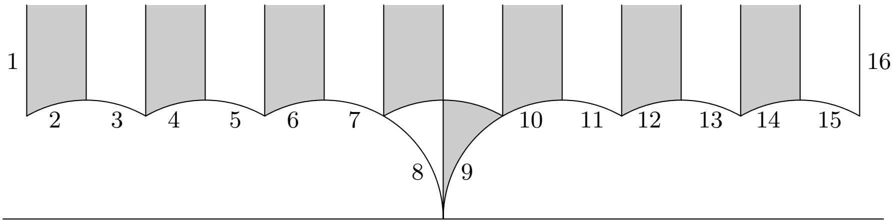


Figure 11 of [Klein 1879a].

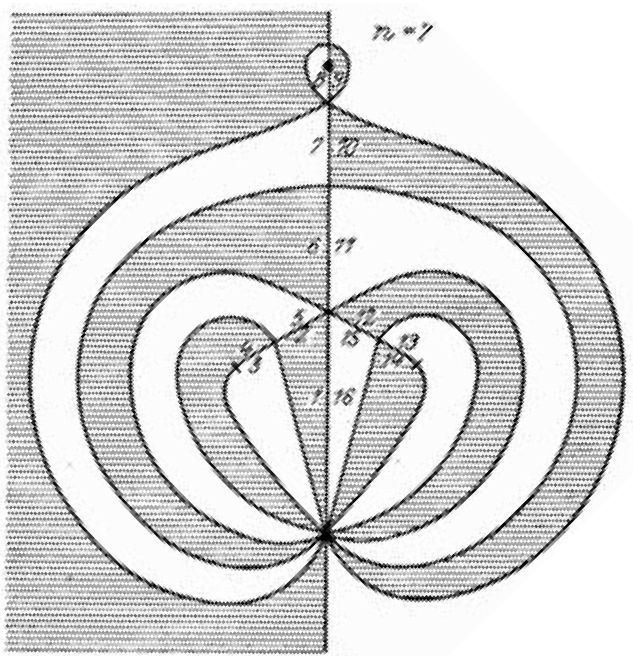


Figure 12 of [Klein 1879a].

[The top figure shows the fundamental polygon in the ω-plane for the modular equation of degree eight. The identification of the sides is given by Klein as follows: $\omega' = \omega+7$ maps 1 to 16, and $\omega' = \omega/(\omega+1)$ maps 6, 7, 8 to 11, 10, 9; these are parabolic transformations. Then $\omega' = (2\omega-7)/(\omega-3)$ maps 15, 14 to 12, 13, and $\omega' = -(2\omega+7)/(\omega+3)$ maps 2, 3 to 5, 4; these are elliptic transformations of period 3. The quotient of the upper half-plane by the group Γ generated by these transformations is a (punctured) sphere, parametrized by the variable τ; the bottom figure shows how the edges of the fundamental polygon become identified in the τ-plane (the figure is combinatorially but not conformally accurate). Thus on the τ-plane there are two order-three branch points of the quotient map (at the lower end of the edges $3 = 4$ and $13 = 14$) and one cusp (at the upper end of the edge $8 = 9$). The Klein surface — the quotient of the ω-plane by the group $\Gamma(7)$ of substitutions congruent to the identity modulo 7 — sits between the ω-plane and the τ-plane: it covers the τ-plane with multiplicity 21, since $\Gamma(7)$ has index 21 in Γ (the quotient $\Gamma/\Gamma(7)$ is the G'_{21} of Sections 2 and 8). –L.]

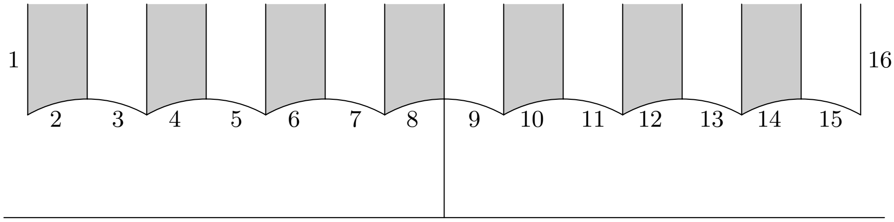


Figure 4 of [Klein 1879b].

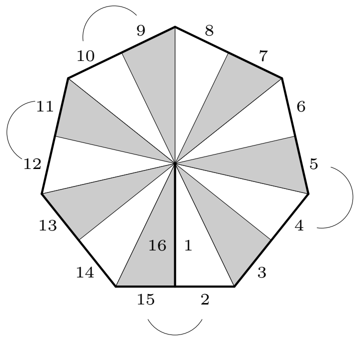


Figure 5 of [Klein 1879b].

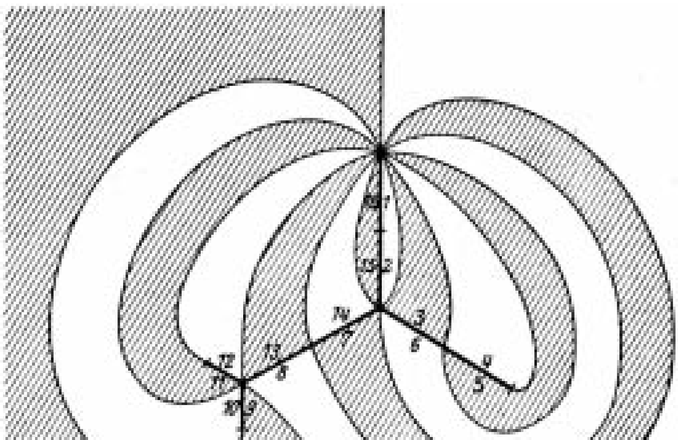


Figure 6 of [Klein 1879b].

[Top: fundamental polygon in the ω-plane for the modular equation of degree seven. Middle: identification of the sides (plus 7, 8 go to 14, 13). Bottom: the $\mathfrak{z}$-plane. –L.]

And conversely: *Every regularly tiled cover defines a particular Galois resolvent with one parameter*. This is because it defines a branching of η with respect to z having the property that each root η_i can be expressed rationally in terms of any other root η_k and of the parameter z.[24]

In the particular case we are considering, there are 168 sheets and three branch points, $J = 0, 1, \infty$. (I will continue to write J instead of z.) At $J = 0$ the sheets are grouped in threes, at $J = 1$ in twos, and at $J = \infty$ in sevens. Thus our surface is covered by $2 \cdot 168$ triangles, which come together in groups at 14 at 24 vertices, in groups of 6 at 56 vertices, and in groups of 4 at 84 vertices. The vertices of these triangles are none other than the a-points, b-points, and c-points of Section 2, and I will maintain this notation here.

From now on we assume that the branch cut in the J-plane is taken to coincide with the real axis. Then the two types of triangles that cover our surface correspond to the two J half-planes, *and the edges of the triangles correspond to real values of* J. I will (as always in the past) shade those triangles that correspond to the upper half-plane ($\operatorname{Im} J > 0$). Thus, for shaded triangles, we have this sequence of vertices:

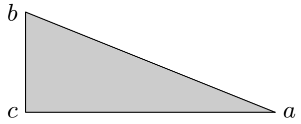


Figure 1.

If we compare this two-triangle surface with the decomposition of the ω-plane into infinitely many triangles from [Klein 1879a] [reproduced at the top of the next page; the labels correspond to the values of J –L.], it is clear that our irrationality J moves over *one* shaded or unshaded triangle when ω traverses a shaded or unshaded triangle, respectively.

Now, the figure at the top of page 315 explained the relation between ω and the root τ of the modular equation of degree eight, and the top and middle figure of page 316 did the same for the root $\mathfrak{z}$ of the modular equation of degree seven. If we move these figures onto our regularly tiled surface and observe that τ and $\mathfrak{z}$ are *rational* functions of $\lambda : \mu : \nu$, so that to any point of our surface there corresponds only one value of τ and one of $\mathfrak{z}$, we obtain the following results:

---

[24] I think it would be a very useful enterprise to list all the regularly tiled covers of low genus p and find out the corresponding equations $F(\eta, z) = 0$. [This problem was solved by W. Dyck in his Inaugural Dissertation [Dyck 1879]; see also the related [Dyck 1880a]. However, there is an error common to those works and the present article: the distinction between *regular* and *regular symmetric* tilings of surfaces had not been yet clearly grasped, and for this reason only the latter type was considered. This error was corrected in [Dyck 1882]; see particularly page 30 and the note therein.

In this connection I would like to stress that Dyck had already devoted a monograph to the study of Riemann surfaces that correspond to *Galois resolvents of modular equations* and achieved a general way to describe them clearly; see [Dyck 1881]. See also [Klein 1923, pp. 166 ff.]. –K.]

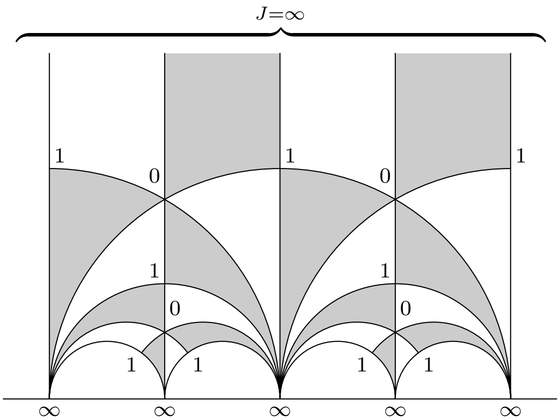


Figure 7 of [Klein 1879a].

Our regularly tiled surface can be divided into 21 *domains such as the one in Figure* 2. *It can also be divided into* 24 *heptagons such as the one in Figure* 3.[25]

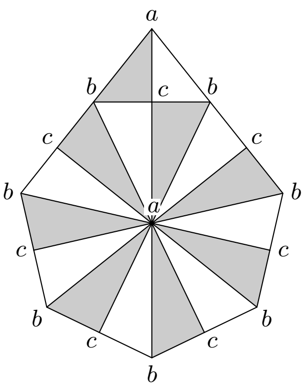


Figure 2.

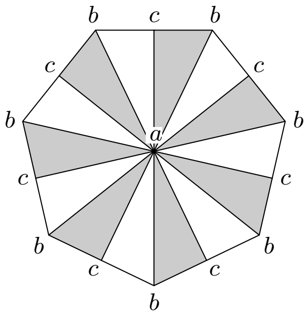


Figure 3.

Figure 2 is divided into two symmetric halves by its middle line. One of the halves is shown on the right. We can therefore say that *our surface is covered with* 42 *alternately congruent and symmetric regions of the type defined by this figure.*[26] I will use this decomposition to develop a completely visual and clear picture of the surface.

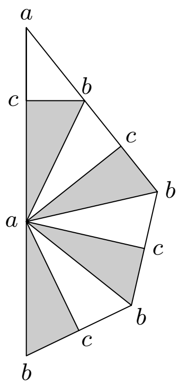


Figure 4.

[25]The grouping into 24 heptagons is analogous to taking the 120 triangles that tile the icosahedron and considering the groups of 10 that surround each of the 12 vertices of the icosahedron.

[26]Thus every region of this type corresponds to the right τ half-plane; see the bottom figure on page 316.

12. Explanation of the Main Figure

The regions just mentioned arrange themselves on our regularly tiled cover in groups of three as follows:

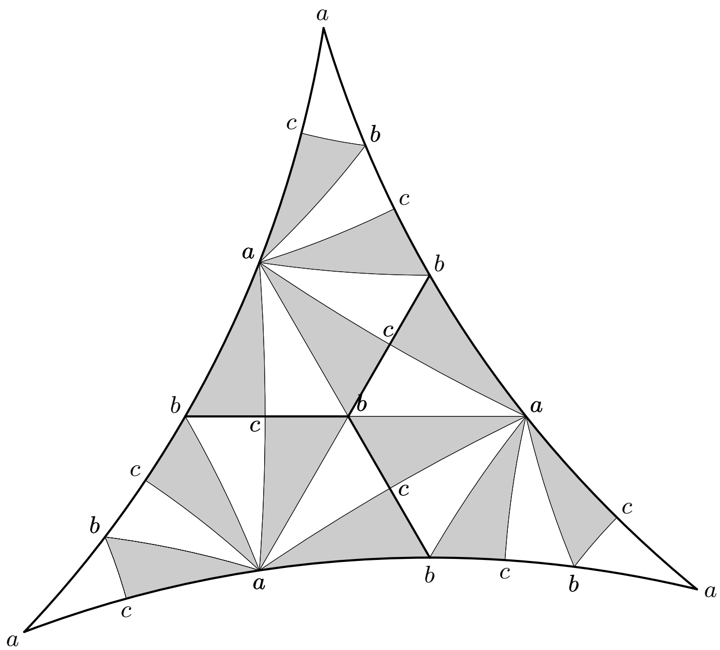


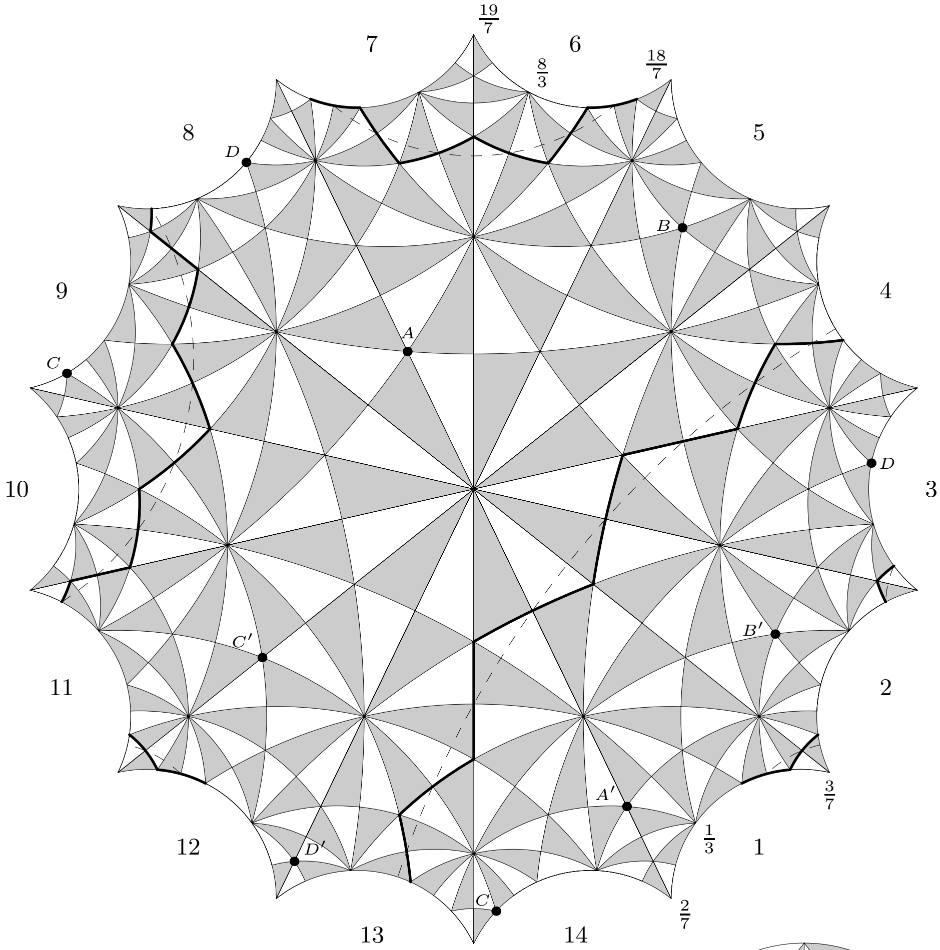
Figure 5.

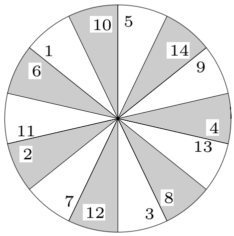
The Main Figure of this article, shown on the next page, is constructed by placing fourteen of these large triangles, with alternate symmetry, around the center point. For the sake of clarity I have made each small triangle out of arcs of circle, having angles of $\pi/7$, $\pi/3$ and $\pi/2$. I now claim that *this figure is a depiction of our regularly tiled surface*, *provided we think of the* 14 *boundary arcs as being identified with each other in the manner stated*.

In fact, our figure contains $2 \cdot 168$ small curved triangles, which exhibit the prescribed behavior at the points where they come together. Starting from this observation, one can look for a suitable correspondence between the boundary arcs, and then carry out the proof that there is no other possible grouping of the $2 \cdot 168$ triangles.

But in order not to make these considerations too abstract, I will resort again to the ω-plane and show on it the same collection of elementary triangles that makes up Figure 5. This is done in Figure 6.

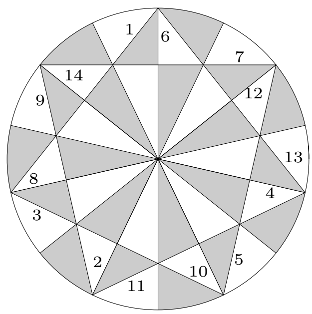
If we now arrange 14 copies of this figure, with alternating symmetry, side by side on the ω-plane, we get the same configuration of triangles shown in the Main Figure. So we must check — and this can be done at once — that this arrangement of 336 triangles in the ω-plane can serve as a *fundamental polygon*





Vertices of one type

Edge identifications:
1 with 6
3 with 8
5 with 10
7 with 12
9 with 14
11 with 2
13 with 4



Vertices of the other type

Main Figure. [Note the three "eightfold ways", discussed in Section 14. Klein's original drawing can be found on page 115. The *Abhandlungen* version is as shown here, differing from the original by a $\pi/7$ rotation. (All the figures were redrawn for the *Abhandlungen*, for the most part with less care; but Figure 5 was improved in that originally the triangles had straight sides and widely different angles, so although combinatorially correct it was harder to grasp than the later version. See also footnote 31, p. 326.) –L.]

for our irrationality; that is, all 168 shaded triangles can be obtained from one of them by means of substitutions

$$\frac{\alpha\omega+\beta}{\gamma\omega+\delta},$$

all of which are distinct modulo 7; and likewise for all 168 unshaded triangles. We must also determine how the edges labeled 1, 2, ..., 14 in the Main Figure match up. Each such edge corresponds on the ω-plane to a *pair* of semicircles meeting the real axis perpendicularly; for example, edge 1 corresponds to a semicircle with endpoints $\omega = \frac{2}{7}, \frac{1}{3}$ and one with endpoints $\omega = \frac{1}{3}, \frac{3}{7}$. Thus, when I claim that edges 1 and 6 match, I must show that the corresponding pairs of semicircles

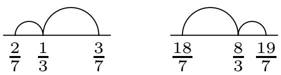


Figure 7.

in the ω-plane are mapped to each other by a substitution that is congruent to the identity modulo 7. This is indeed the case: the substitution

$$\omega' = \frac{113\omega - 35}{42\omega - 13}$$

maps $\frac{2}{7}$ to $\frac{19}{7}$ and $\frac{1}{3}$ to $\frac{8}{3}$, and so maps the semicircle that meets the real axis at $\frac{2}{7}$ and $\frac{1}{3}$ to the semicircle that meets the real axis at $\frac{19}{7}$ and $\frac{8}{3}$.

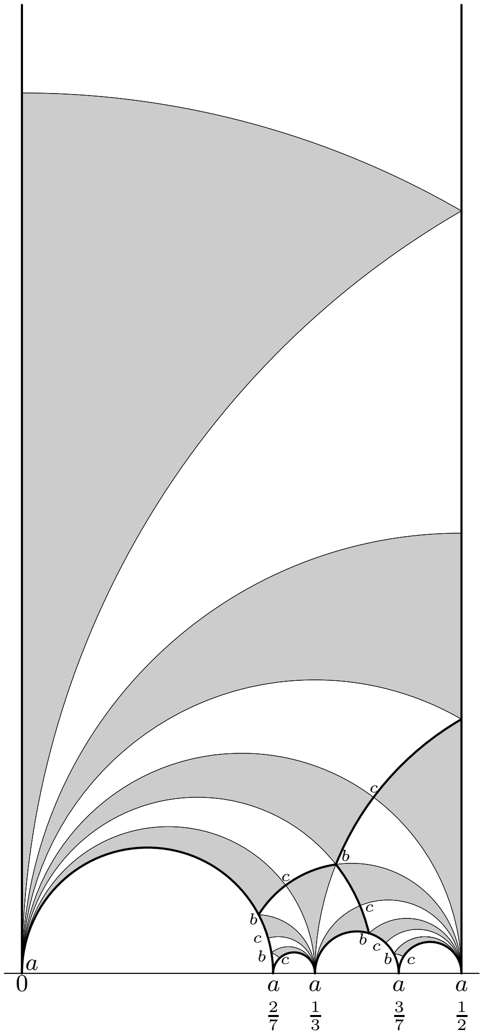


Figure 6.

Similarly, the substitution

$$\omega' = \frac{55\omega - 21}{21\omega - 8}$$

maps $\frac{1}{3}$ to $\frac{8}{3}$ and $\frac{3}{7}$ to $\frac{18}{7}$, which shows that the second halves of the pairs of semicircles match. I have marked the points $\frac{2}{7}$, $\frac{1}{3}$, $\frac{3}{7}$ and $\frac{18}{7}$, $\frac{8}{3}$, $\frac{19}{7}$ at the corresponding places in the Main Figure. *This argument shows that edges* 1 *and* 6 *are to be identified in such a way that* $\frac{2}{7}$ *coincides with* $\frac{19}{7}$ *and* $\frac{3}{7}$ *with* $\frac{18}{7}$.

In a similar way[27] one finds that *the following pairs of edges are to be identified*:

1 *with* 6, 3 *with* 8, 5 *with* 10, 7 *with* 12, 9 *with* 14, 11 *with* 2, 13 *with* 4,

and in each case vertices of the same type come together. What I mean by "vertices of the same type" is made clear by the figure; the small circles below the Main Figure illustrate how the 14 vertices come together according to their type, forming two a points.

If one were to actually bend the polygon of the Main Figure and glue the edges together, the result would be a very confusing figure. It is better to remain on the plane and complement the Main Figure with the edge identifications and the two small figures showing the incidence at the vertices. In this way one reaches the results compiled in the next section.

13. The 28 Symmetry Lines

By a *symmetry line* of our covering surface I will mean a line made up of triangle edges and not having kinks anywhere — going straight, so to speak, through a, b, and c-points. The surface is indeed symmetric with respect to such lines: as an example of a symmetry line we can take the vertical center line of the Main Figure, so long as we make it into a closed curve by adding edge 5, or, equivalently, edge 10; these two edges are symmetrically placed with respect to the center line, and moreover the gluing scheme for the remaining edges is symmetric with respect to this line.

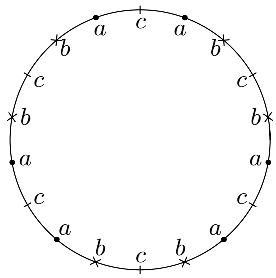


Figure 8.

This example also shows that such a symmetry line must contain six points of each type a, b, c, in the sequence indicated in Figure 8.

Next we have, most importantly:

There are 28 *symmetry lines*. Together they comprise all the triangle edges, and so they exhaust the points on the surface that correspond to *real* values of J.

These symmetry lines are, for many purposes, the easiest means of orientation on our surface; I will use them here to characterize the groupings of a, b,

[27] [Each point $k+\frac{1}{3}$, for $k=0,1,\dots,6$, lies on the edge labeled $2k+1$ on the Main Figure, and each point $k'-\frac{1}{3}$, for $k'=1,2,\dots,7$, lies on the edge $2k$. When is a point $k+\frac{1}{3}$ equivalent to a point $k'-\frac{1}{3}$ under our group? The condition

$$k+\tfrac{1}{3}=\frac{(7a+1)(k'-\frac{1}{3})+7b}{7c(k'-\frac{1}{3})+(7d+1)}$$

yields, when considered modulo 7,

$$k'-k\equiv\tfrac{2}{3}\equiv 3 \pmod 7, \quad \text{or} \quad 2k'-(2k+1)\equiv 5 \pmod 7;$$

that is, edges $2k+1$ and $2k+6$ are to be identified. –B.-H.]

and c points. Then it is easy to form an idea of the corresponding one-to-one transformations of our surface into itself.

The 7 *symmetry lines that meet at an a-point also meet at the other two points in the same triple* (coming from a G'_{21}: see page 294). An example of such a triple is given by the center of our figure together with the two a-points coming from the two kinds of vertices along the boundary (seven of each kind). The process for passing from the closed regularly tiled surface to the Main Figure can be described as follows: Choose on the surface two out of a triple of a-points and cut along the seven pieces of symmetry lines that go from one of these a-points to the other. Since the surface has genus $p = 3$, the result is simply connected and has one boundary curve, and when stretched out on the plane, it becomes our Main Figure of page 320.

Clearly, any *two* triples of a-points determine exactly *one* symmetry line, on which the points of the two triples alternate.

The 3 *symmetry lines that meet at a b-point also meet at the other b-point with which it forms a pair.* Examples of pairs of b-points are given in the figure by A, A'; B, B'; C, C'; D, D'; we will return to them later. To each such triple of symmetry lines, and so to each pair of b-points, is associated a symmetry line, characterized by the fact that it intersects the lines of the triple in two c-points. This gives a one-to-one correspondence between the 28 pairs of b-points and the 28 symmetry lines.

The 2 *symmetry lines that meet at a c-point meet again at another c-point. There are two more symmetry lines that do not intersect the first two and that meet each other at another pair of c-points. In this way one obtains the quadruples of c-points.*

14. Definitive Shape of Our Surface

The more regular a figure is, the more it tends to be intuitive and easy to grasp. Thus I would like to put our regularly tiled surface into a shape that allows as many as possible of the 168 one-to-one transformations to be realized as *rotations*. Now, we know all the finite groups that can be realized by rotations: they correspond to the regular polyhedra. There is no group of 168 rotations [in three-dimensional space] in the sense we are talking about. On the other hand, we have already remarked in Section 1 that the 24 substitutions in a G''_{24} stand in the same relation to one another as the rotations that take an octahedron to itself. This suggests that it may be possible *to give our surface such a shape that it is sent to itself by the rotations of an octahedron.*

For this purpose we must first find four b-points that are permuted by the substitutions of a G''_{24}. This can be accomplished easily if we group the 14 triangles that meet at each a-point, making 24 heptagons that together cover the whole surface, as discussed earlier. *Then there are* $2 \cdot 7$ *ways to choose four pairs of b-points so that all* 24 *heptagons have one of the chosen b-points as a* *vertex*.[28] The four point pairs A, A', B, B', C, C', D, D', already mentioned, form such a quadruple. Six more are obtained by rotating the figure around the center in multiples of $2\pi/7$, and the remaining seven by reflecting the first seven in any symmetry line, say the vertical center line.

Now cut the surface (after having glued the three zigzag paths shown in the Main Figure as thick lines weaving around the dashed curves. The result is a sextuply connected surface with six boundary curves [a sphere minus six disks –L.], *and this surface can be stretched symmetrically onto a sphere in such a way that the eight points A, A', etc. coincide with the vertices of an inscribed cube*, *the vertices of the dual octahedron remain uncovered*, *and the twelve midpoints of the spherical octahedron's edges coincide with the c-points of the surface*. For greatest clarity I have sketched a drawing showing only one of the octants of the sphere (see Figure 9).

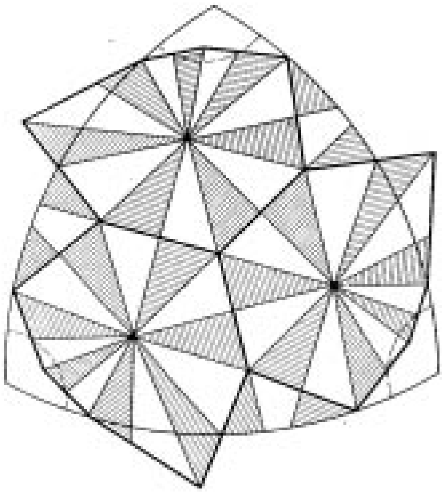


Figure 9.

The three heptagons that meet at the center of the octant fall partly outside the octant. But since this is true also about the heptagons that cover the neighboring octants, the only part of the octant that is not covered by the surface is the corners.

To obtain an image of the surface as a whole we must know how the boundary curves that surround the corners of the octahedron are to be joined together. The answer can be read off by comparing with the earlier figure, and it is very simple: *each point must be identified with the diametrically opposed point*.

These identifications can be carried out without breaking the desired octahedral symmetry: *one just has to bring together the boundary curves through infinity in such a way that the intersection with the plane at infinity consists of the curves shown in the Main Figure and in Figure* 9 *as dashed lines*. Therefore the heptagons that spread out from the center of the octant reach out in part beyond infinity, so that a total of twelve c-points lie on the plane at infinity. So the surface itself goes out to infinity in much the same way as the union of three congruent hyperboloids of rotation whose axes meet at right angles.[29] [See Figure 8 of [Gray 1982], page 127 in this volume. –L.]

[28] Also, the existence of the resolvent of degree eight can easily be proved using these heptagons: *There are eight ways to choose three heptagons so that the remaining 21 heptagons are adjacent to one of the three.*

[29] [Dyck prepared at the time a nice model of the surface in this form, for the Mathematics Institute of the Technische Hochschule München. –K.]

If one wishes to check that the 24 transformations expressible by rotations of the octahedron fix the number of points asserted earlier, one should keep in mind that *a rotation of period two fixes not only points on the rotation axis but also points on the line at infinity that lies perperdicular to the rotation axis.* Our surface is not intersected by the diagonals of the octahedron, but it is intersected four times by each line at infinity perpendicular to a diagonal of the octahedron. The diameters going through the midpoints of the edges of the octahedron intersect the surface twice, as do the lines at infinity perpendicular to these diameters. Finally, the diagonals of the cube have exactly two intersections with the surface. Therefore a rotation of period four fixes *no points*, one of period two fixes *four points*, and one of period three fixes *two points.* This is all as it should be.

15. The Real Points of the Curve of Order Four

I would like to conclude by showing how these relative positions stand out when we consider *the real points of the order-four curve*

$$\lambda^3\mu + \mu^3\nu + \nu^3\lambda = 0.$$

The coordinate triangle may be taken to be equilateral, the coordinates being proportional to the distance to the sides. Then the bitangent $\lambda+\mu+\nu = 0$ is the line at infinity; its contact points $1:\alpha:\alpha^2$ and $1:\alpha^2:\alpha$ are the two cyclic points. The line at infinity is therefore an *isolated* bitangent. The six collineations of the corresponding G_6' are the only real ones among the 168; they consist of the three rotations through 120 degrees about the center of the coordinate triangle and of the reflections in three lines going through this same center. These lines are the only three real perspective axes; the related perspective centers lie at infinity, orthogonally to the lines. From the inflection triangle $\lambda\mu\nu = 0$, the three reflections give rise to a second real inflection triangle $\lambda'\mu'\nu' = 0$.

We now consider form (9) of the curve's equation:

$$49y_1(y_1 + y_2 + y_3)(y_1 + \alpha y_2 + \alpha^2 y_3)(y_1 + \alpha^2 y_2 + \alpha y_3) \; - \; 3(4y_1^2 - 7y_2y_3)^2 = 0.$$

We replace y_1 by 1 (since $y_1 = 0$ is the line at infinity), and replace y_2, y_3 by $x + iy$, $x - iy$, since the corresponding axes go through the cyclic points. We obtain

$$49(2x + 1)(-x + \sqrt{3}\,y + 1)(-x - \sqrt{3}\,y + 1) - 3(4 - 7(x^2 + y^2))^2 = 0.$$

The bitangents

$$2x + 1 = 0, \quad -x + \sqrt{3}\,y + 1, \quad -x - \sqrt{3}\,y + 1 = 0$$

again form an equilateral triangle, of altitude $\frac{3}{2}$ and side length $\sqrt{3}$. Its intersection with the circle of radius $2/\sqrt{7}$ around the center consists of contact points of bitangents, *so we have three nonisolated bitangents*. One can check that all other

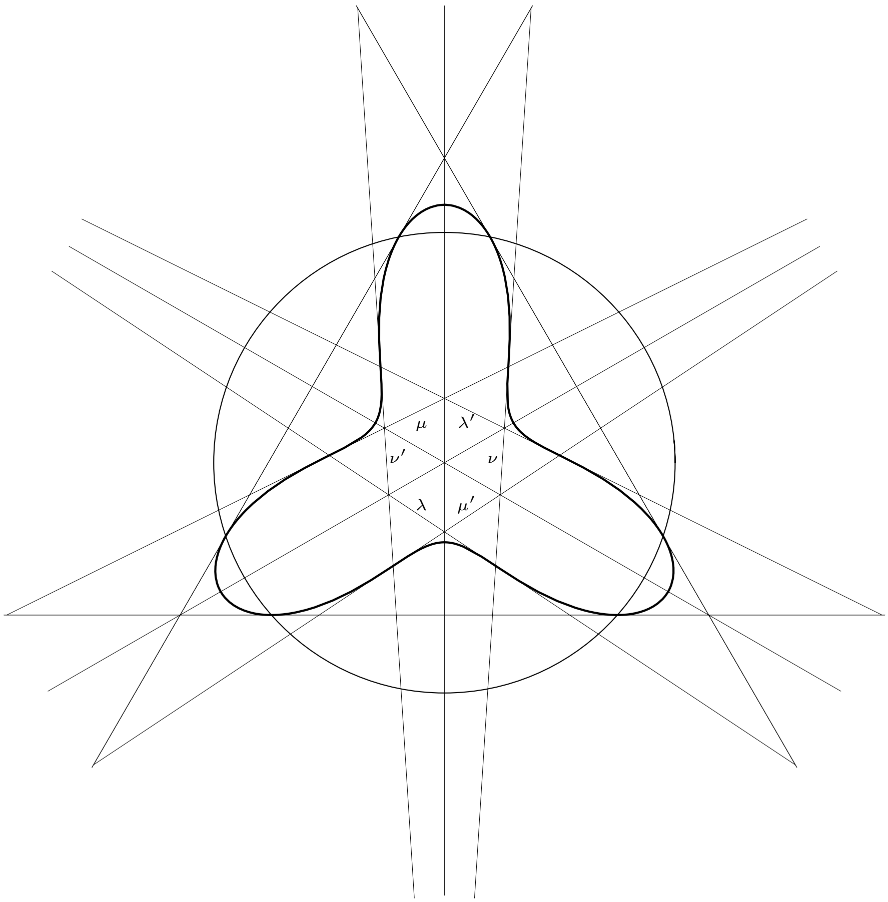


Figure 10.

bitangents are imaginary. *Thus our curve has one branch*[30] *and is inscribed in the triangle of nonisolated bitangents.* The accompanying diagram[31] (Figure 10) shows, in addition to the bitangents in question, the circle through the contact points, the three real perspective axes, and the two real inflection triangles.

The real curve so obtained has a very simple interpretation in terms of the Riemann surface: *it represents one of the* 28 *symmetry lines.* Indeed, real values of λ, μ, ν yield real values of J, and the symmetry lines are characterized by J real.

---

[30] See [Zeuthen 1874].

[31] [The schematic illustration in the original was replaced in the reprint by a figure precisely computed by Haskell, which appears in his Göttingen dissertation [Haskell 1891]. In this work, done at my instigation, Haskell applies the ideas developed in [Klein 1874; 1876] to the curve $\lambda^3\mu + \mu^3\nu + \nu^3\lambda = 0$ and clarifies the results with illustrations. –K.]

This symmetry line is clearly associated with the isolated bitangent at infinity and so, like the bitangent, it is sent to itself by the substitutions of a G_6'. It contains six each of the a, b, and c-points, and it can be seen in Figure 10 that these points follow one another in the sequence expected for a symmetry line (Figure 8).

Munich, early November 1878.

[Additional Remarks Concerning Some of the Literature]

[The mathematical literature concerning the fourth-order curve $\lambda^3\mu + \mu^3\nu + \nu^3\lambda$ and the thus defined algebraic structure having 168 one-to-one transformations has multiplied since the publication of this article, particularly in its geometric aspects. It is not possible to cover it in detail here, but I will at least mention some highlights.

In what concerns the algebraic side of the question, we refer to Gordan's extensive investigations, discussed in [Klein 1922, pp. 426 ff.]. Here I will add a discussion of the role played by n-th roots of unit, which come up in the articles reprinted in the first half of [Klein 1923] and also in [Klein 1922]. When one considers Galois problems, these are "natural" irrationalities: for example, the fifth root of unity can be represented, by virtue of the icosahedral substitutions, as a quotient of appropriately chosen roots of the icosahedral equation. The same is true of the partition equations of elliptic functions, as a consequence of the so-called "Abel relations". See [Klein 1885, footnote 37]. For the modular equations of the functions $J(\omega)$, however, the n-th roots of unity are no longer "natural", but the Gaussian sum $\sqrt{(-1)^{(n-2)/2}n}$ formed from them is. See, for example, [Fricke 1922, p. 462]. This is also true of the special resolvents of fifth, seventh, and eleventh degree, treated in [Klein 1879b; 1879d]. (Cf. for instance [Fricke 1922, p. 482].) These results are important in order to determine in individual cases not only the monodromy group,[32] to which the exposition in the text has limited itself, but also the Galois group, taking as a basis the domain of rationality of the rational numbers.

Another line of research concerns the three globally finite integrals of our fourth-order curve. It seems particularly remarkable that their periods can be explicitly given. Poincaré [1883] and Hurwitz [1886, p. 123] find, for an appropriate choice of crosscuts, the period matrix

$$\begin{array}{cccccc} 1 & 0 & 0 & \tau & \tau-1 & -\tau \\ 0 & 1 & 0 & \tau-1 & -\tau & \tau \\ 1 & 0 & 0 & -\tau & \tau & \tau-1 \end{array}$$

[32] [The concept of monodromy group was introduced by Hermite in [Hermite 1851]. The name appears for the first time, so far as I know, in [Jordan 1870, p. 278]. –K.]

where τ denotes the quadratic irrational number $\frac{1}{4}(1 + i\sqrt{7})$. This implies, in particular, that our Riemann surface has a multiple cover by an elliptic surface of singular modulus $\omega = \frac{1}{2}(-1+i\sqrt{7}) = (\tau-1)/\tau$, and therefore having the rational invariant $J(\omega) = -5^3/2^6$. Moreover Hurwitz [1885] has studied the integral of first type as a function of ω and in the coefficients of its power series development in $q^2 = e^{2\pi i\omega}$ he found those number-theoretic functions that Gierster ran into in the construction of class number relations of rank seven. See [Klein 1923, p. 5]. More details on the subject can be found in the "Modular functions".

Perhaps our curve achieves the greatest prominence in that the Main Figure on page 320, when placed inside a disk whose boundary is orthogonal to its arcs, provides the first concrete example of uniformization of an algebraic curve of higher genus. For this reason it became for me the best prop in building the general uniformization results in [Klein 1882a; 1882b; 1883].

The considerations in the text find an immediate continuation in a note by Dyck [1880b] about the normal curve $\lambda^4 + \mu^4 + \nu^4$ pertaining to the main congruence group of rank eight and admitting 96 one-to-one transformations onto itself, and particularly in Fricke's investigations about the ternary Valentiner–Wiman group and the transformation theory of triangle functions for a triangle with angles $\pi/5$, $\pi/2$, $\pi/4$. (Published as an appendix in [Fricke and Klein 1912]. See also [Klein 1922, pp. 501–502].) –K.]

References

[Brill and Nöther 1874] A. Brill and M. Nöther, "Ueber die algebraischen Functionen und ihre Anwendung in der Geometrie", *Math. Annalen* **7** (1874), 269–310.

[Brioschi 1868] F. Brioschi, "Sopra le equazioni dell'ottavo grado che hanno lo stesso gruppo delle equazioni del moltiplicatore corrispondente alla trasformazione di settimo ordine delle funzioni ellittiche", *Rendiconti dell' Istituto Lombardo* (2) **1** (1868), 68–72. Reprinted as [Brioschi 1901–09, vol. 3, CXVII, pp. 243–247].

[Brioschi 1878/79] F. Brioschi, "Sopra una classe di equazioni modulari", *Annali di Matematica pura ed applicata* (2) **9** (1878/79), 167–172. Reprinted as [Brioschi 1901–09, vol. 2, LXXV, pp. 193–198].

[Brioschi 1879] F. Brioschi, "Ueber die Jacobi'sche Modulargleichungen vom achten Grad", *Math. Annalen* **15** (1879), 241–250. Reprinted as [Brioschi 1901–09, vol. 5, CCXXXV, pp. 225–237].

[Brioschi 1901–09] F. Brioschi, *Opere matematische*, edited by G. Ascoli et al., U. Hoepli, Milano, 1901–09.

[Clebsch 1876] A. Clebsch, *Vorlesungen über Geometrie, 1: Geometrie der Ebene*, edited by F. Lindemann, Teubner, Leipzig, 1876. Reprinted by Johnson Reprint Corp., New York, 1968.

[Clebsch and Gordan 1866] A. Clebsch and P. Gordan, *Theorie der Abelschen Functionen*, Teubner, Leipzig, 1866. Reprinted by Physica-Verlag, Wurzburg, 1967.

[Dyck 1879] W. von Dyck, *Über regulär verzweigte Riemannsche Flächen und die durch sie definierten Irrationalitäten*, Inauguraldissertation, Ludwig-Maximilians-Universität, München, 1879.

[Dyck 1880a] W. von Dyck, "Ueber Untersuchung und Aufstellung von Gruppe und Irrationalität regulärer Riemann'scher Flächen", *Math. Annalen* **17** (1880), 473–509.

[Dyck 1880b] W. von Dyck, "Notiz über eine reguläre Riemann'sche Fläche vom Geschlechte drei und die zugehörige 'Normalcurve' vierter Ordnung", *Math. Annalen* **17** (1880), 510–516.

[Dyck 1881] W. von Dyck, "Versuch einer übersichtlichen Darstellung der Riemann'schen Fläche, welche der Galois'schen Resolvente der Modulargleichung für Primzahltransformation der elliptischen Functionen entspricht", *Math. Annalen* **18** (1881), 507–527.

[Dyck 1882] W. von Dyck, "Gruppentheoretische Studien", *Math. Annalen* **20** (1882), 1–44.

[Fricke 1922] R. Fricke, *Die elliptischen Funktionen und ihre Anwendungen,* 2*: Die algebraischen Ausführungen*, Teubner, Leipzig, 1922. Reprinted by Johnson Reprint Corp., New York, 1972.

[Fricke and Klein 1912] R. Fricke and F. Klein, *Vorlesungen über die theorie der automorphen Functionen,* 2*: die functionentheoretischen Ausführungen und die Anwendungen*, Teubner, Leipzig, 1912. Reprinted by Johnson Reprint, New York, and Teubner, Stuttgart, 1965.

[Gierster 1881] J. Gierster, "Die Untergruppen der Galois'schen Gruppe der Modulargleichungen für den Fall eines primzahligen Transformationsgrades", *Math. Annalen* **18** (1881), 319–365.

[Gray 1982] J. Gray, "From the history of a simple group", *Math. Intelligencer* **4**:2 (1982), 59–67. Reprinted in this collection.

[Halphen 1884] G.-H. Halphen, "Sur une équation différentielle linéaire du troisième ordre", *Math. Annalen* **24** (1884), 461–464.

[Haskell 1891] M. W. Haskell, "Ueber die zu der Curve $\lambda^3\mu + \mu^3\nu + \nu^3\lambda = 0$ im projectiven Sinne gehörende mehrfache Ueberdeckung der Ebene", *Amer. J. Math.* **13** (1891), 1–52.

[Hermite 1851] C. Hermite, "Sur les fonctions algébriques", *C. R. Acad. Sci.* (Paris) **32** (1851), 458–461. Reprinted as pp. 276–280 in his *Œuvres*, vol. 1, Gauthier-Villars, Paris, 1905.

[Hesse 1855] L. O. Hesse, "Ueber Determinanten und ihre Anwendung in der Geometrie, insbesondere auf Curven vierter Ordnung", *J. für reine angew. Math.* **49** (1855), 243–264. Reprinted as #24, pp. 319–343 his *Gesammelte Werke*, Chelsea, New York, 1972.

[Hurwitz 1885] A. Hurwitz, "Ueber Relationen zwischen Classenanzahlen binärer quadratischer Formen von negativer Determinante", *Math. Annalen* **25** (1885), 157–196.

[Hurwitz 1886] A. Hurwitz, "Ueber einige besondere homogene lineare Differentialgleichungen", *Math. Annalen* **26** (1886), 117–126.

[Jordan 1870] M. C. Jordan, *Traité des substitutions et des équations algébriques*, Gauthier-Villars, Paris, 1870.

[Jordan 1878] M. C. Jordan, “Mémoire sur les équations différentielles linéaires à intégrale algébrique”, *J. für reine angew. Math.* **84** (1878), 89–215. Reprinted as pp. 13–140 of his *Œuvres*, vol. 2, Gauthier-Villars, Paris, 1962.

[Jung and Armenante 1869] G. Yung and A. Armenante, “Sopra una equazione dell'8^o grado”, *Giornale di Matematiche ad uso degli studenti delle università italiane* **7** (1869), 98–104.

[Klein 1874] F. Klein, “Ueber eine neue Art der Riemann'schen Flächen (erste Mitteilung)”, *Math. Annalen* **7** (1874), 558–566. Reprinted as [Klein 1922, XXXVIII, pp. 89–98].

[Klein 1876] F. Klein, “Ueber eine neue Art der Riemann'schen Flächen (zweite Mitteilung)”, *Math. Annalen* **10** (1876), 398–416. Reprinted as [Klein 1922, XL, pp. 136-155].

[Klein 1878a] F. Klein, “Ueber Gleichungen siebenten Grades (erste Mitteilung)”, *Sitzungsberichte der physikalisch-medizinischen Sozietät zu Erlangen* **10** (1878), 110–111. Reprinted as [Klein 1922, LVI, pp. 388–389].

[Klein 1878b] F. Klein, “Ueber Gleichungen siebenten Grades (zweite Mitteilung)”, *Sitzungsberichte der physikalisch-medizinischen Sozietät zu Erlangen* **10** (1878), 119–123.

[Klein 1879a] F. Klein, “Ueber die Transformation der elliptischen Funktionen und die Auflösung der Gleichungen fünften Grades”, *Math. Annalen* **14** (1879), 111–172. Reprinted as [Klein 1923, LXXXII, pp. 13–75].

[Klein 1879b] F. Klein, “Ueber die Erniedrigung der Modulargleichungen”, *Math. Annalen* **14** (1879), 417–427. Reprinted as [Klein 1923, LXXXIII, pp. 76–89].

[Klein 1879c] F. Klein, “Ueber die Auflösung gewisser Gleichungen vom siebenten und achten Grade”, *Math. Annalen* **15** (1879), 251–282. Reprinted as [Klein 1922, LVII, pp. 390–438].

[Klein 1879d] F. Klein, “Ueber die Transformation elfter Ordnung der elliptischen Modulfunktionen”, *Math. Annalen* **15** (1879), 533–555. Reprinted as [Klein 1923, LXXXVI, pp. 140–168].

[Klein 1882a] F. Klein, “Ueber eindeutige Funktionen mit linearen Transformationen in sich (erste Mitteilung)”, *Math. Annalen* **19** (1882), 565–568. Reprinted as [Klein 1923, CI, pp. 622–626].

[Klein 1882b] F. Klein, “Ueber eindeutige Funktionen mit linearen Transformationen in sich (zweite Mitteilung)”, *Math. Annalen* **20** (1882), 49–51. Reprinted as [Klein 1923, CII, pp. 627–629].

[Klein 1883] F. Klein, “Neue Beiträge zur Riemann'schen Funktionentheorie”, *Math. Annalen* **21** (1883), 141–218. Reprinted as [Klein 1923, CIII, pp. 630–710].

[Klein 1885] F. Klein, “Ueber die elliptischen Normalkurven der n-ten Ordnung”, *Abhandlungen der mathematischen-physikalischen Klasse der Sächsischen Kgl. Gesellschaft der Wissenschaften* **13**:4 (1885). Reprinted as [Klein 1923, XC, pp. 198–254].

[Klein 1922] F. Klein, *Gesammelte Mathematische Abhandlungen, 2: Anschauliche Geometrie; Substitutionsgruppen und Gleichungstheorie; zur mathematischen Physik*, edited by R. Fricke and H. Vermeil, Springer, Berlin, 1922. Reprinted by Springer, 1973.

[Klein 1923] F. Klein, *Gesammelte Mathematische Abhandlungen, 3: Elliptische Funktionen* etc., edited by R. Fricke et al., Springer, Berlin, 1923. Reprinted by Springer, 1973.

[Poincaré 1883] H. Poincaré, "Sur l'intégration algébrique des équations linéaires", *C. R. Acad. Sci.* (Paris) **97** (1883), 1189–1191. Reprinted as pp. 103–105 in his *Œuvres*, vol. 3, Gauthier-Villars, Paris, 1934.

[Riemann 1861/62] B. Riemann, "Vorlesungen über die allgemeine Theorie der Integrale algebraischer Differentialien", pp. 1–66 in *Bernard Riemann's gesammelte Mathematische Werke: Nachträge*, edited by M. Noether and W. Wirtinger, Teubner, Leipzig, 1902. Lectures given in the winter semester 1861/62. Reprinted as pp. 599–664 in the one-volume edition of Riemann's *Gesammelte Mathematische Werke*, Springer and Teubner, 1990.

[Serret 1866] J.-A. Serret, *Cours d'algèbre supérieure*, 3rd ed., Gauthier-Villars, Paris, 1866.

[Weber 1876] H. Weber, *Theorie der Abel'schen Functionen vom Geschlecht* 3, G. Reimer, Berlin, 1876.

[Zeuthen 1874] H. G. Zeuthen, "Sur les différentes formes des courbes plane du quatrième ordre", *Math. Annalen* **7** (1874), 410–432.

Felix Klein (1849–1925)

Silvio Levy
Mathematical Sciences Research Institute
1000 Centennial Drive
Berkeley, CA 94720-5070
United States